

REPRESENTING OURSELVES AS RATIONAL AGENTS

by

Matthew D. Mosdell

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

The University of Utah

December 2012

Copyright © Matthew D. Mosdell 2012

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Matthew D. Mosdell
has been approved by the following supervisory committee members:

<u>Elijah Millgram</u>	, Chair	<u>Oct. 15, 2012</u> Date Approved
<u>Chrisoula Andreou</u>	, Member	<u>Oct. 15, 2012</u> Date Approved
<u>Eric Hutton</u>	, Member	<u>Oct. 15, 2012</u> Date Approved
<u>Ronald Mallon</u>	, Member	<u>Oct. 15, 2012</u> Date Approved
<u>Mark Wilson</u>	, Member	<u>Oct. 15, 2012</u> Date Approved

and by Stephen Downes, Chair of
the Department of Philosophy

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

It is often assumed that rational agents are unified agents. As a normative feature of agency, the idea that fully rational creatures are, in a certain sense, unified, suggests that understanding ourselves as rational creatures requires understanding our reasons for acting as reasons that can be represented with a single account of agency. In this dissertation, I argue against such a view by showing that features of our own agency preclude constructing such a representation. To be the type of creatures we are, we have to act in ways that cannot be represented with a single, unified theory of rational agency. Instead, making sense of ourselves as rational creatures requires a number of different models that cannot be nicely fitted together. The upshot is that a unified account of rational agency may not be possible.

For My Father

CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	vii
CHAPTERS	
1. INTRODUCTION	1
1.1 Representing Agency	1
1.2 Models and Theories	3
1.3 Elements of Agency	5
1.3.1 Instrumentalism	6
1.3.2 Emotions	8
1.3.3 Practical Learning	11
1.3.4 Autonomy	13
1.4 Is Agency Worth Modeling?	14
1.4.1 Practical Reasoning	14
1.4.2 Ethics	15
1.4.3 Social and Political Philosophy	16
1.5 Why the Modeling Approach?	17
1.6 Prospectus	20
2. INSTRUMENTAL FACADES	22
2.1 Introduction	22
2.2 Three Views of Instrumental Agency	24
2.2.1 The Rational Ideal	25
2.2.2 Autonomous Agents	27
2.2.3 Planning Agents	30
2.3 Prescriptive Discord	33
2.3.1 Smithian Agents	34
2.3.2 Deep Desires	35
2.3.3 Planning Agents	37
2.4 Nonoptional Capacities of Agency	38
2.4.1 Subjunctive Thought	39
2.4.2 Self-Intelligibility	41
2.4.3 Temporal Organization	44
2.4.4 Nonoptional Views	46
2.5 Models of Agency	47
2.5.1 Against Integration	47
2.5.2 Facades of Representation	52
2.5.3 Disunified Models	54
2.6 The Disunity of Instrumentalism	55

3. MODELING EXPRESSIVE ACTIONS	57
3.1 Introduction	57
3.2 Describing Intentional Action	58
3.3 Vogler On Acts of Expression	62
3.3.1 The Challenge of Expressive Acts	63
3.3.2 Art as Expressive Action	64
3.4 Befittingly Artistic	66
3.4.1 Against Befitting Explanations	69
3.4.2 A Dilemma	72
3.4.3 Emotional Expression	73
3.5 Toward an Alternative	76
3.5.1 Ordering Emotion	77
3.5.2 Emotion and Thought	78
3.6 Conclusion	80
4. ACQUIRING KNOW HOW	82
4.1 Introduction	82
4.2 Practical Intellectualism	84
4.3 Intellectualizing Knowledge-How	86
4.3.1 Acquisition	88
4.3.2 Modes of Thought	89
4.3.3 Impending Regress	93
4.3.4 Learning to Act	95
4.4 Basic Powers and Intelligent Actions	97
4.5 Intellectualized Action	100
4.5.1 Anscombian Action	101
4.5.2 Anscombe Updated	103
4.5.3 Learning to Act	109
4.5.4 Coordinated Movement	112
4.6 Modeling Explorative Actions	112
4.7 Conclusion	115
5. CONCLUSION	117
5.1 Introduction	117
5.2 The Role of Unity	117
5.3 A Sketch of Disunity	121
5.4 Models of Agency	124
REFERENCES	127

ACKNOWLEDGMENTS

I would first like to acknowledge and thank Lije Millgram as my advisor. No one has taught me more or challenged me to think harder than Lije, and although it has at times been maddening, I am deeply grateful. My understanding, appreciation, and love of philosophy would not be what it is without Lije. Thank you!

I also want to thank my committee members and those who have read, commented on, or otherwise been involved with earlier portions of this dissertation. Chrisoula Andreou, Eric Hutton, Ron Mallon, Monika Piotrowska, Jim Tabery, Mike Wilson, and Mark Wilson have all, in one way or another, been very helpful. I am also indebted to everyone who has over the years read, commented on, and criticized my work as part of attending the semiweekly kaffeeklatsch. My dissertation would be quite different if not for long, influential discussions with my fellow graduate students. I am especially grateful for discussions with Meg Bowman and Matt Berk. To the rest of the philosophy department at the University of Utah, thank you.

Finally, I am grateful to the Marriner S. Eccles Graduate Fellowship in Political Economy, and to the Department of Philosophy at the University of Utah for financial support. It is only because of their support that my life as a graduate student has been financially possible.

CHAPTER 1

INTRODUCTION

1.1 Representing Agency

To explain the natural order of things, it has been the habit of individuals to build models. Ptolemy proposed a model of the universe to explain the movement of the stars; Niels Bohr designed a model of the hydrogen atom to explain its associated emission wavelength; James Watson and Francis Crick built the double helix model of DNA to better understand how it functions as an information transfer mechanism; and scientists, economists, engineers, and so on continue to build models in order to investigate and understand a variety of phenomena.

The advantage of models is that they serve to simplify complexity, making it easier to explore systems that are too messy or too complicated to be fully understood intact and unmanipulated. This advantage is important when modeling features of the human system. Consider, for example, the color-coded models used to show the size, function, and arrangement of blood vessels constitutive of the human circulatory system. Without such models, it would be too messy and too complicated for anatomy students to explore that aspect of human physiology. Of course, the physical features of the human system are not the only ones we aim to understand. In addition to the way our physical parts fit together to form a functioning whole, we also want to understand the way in which human beings function as rational agents—or as creatures capable of producing actions for reasons. As with human physiology, investigating human beings as rational agents can be a complex, messy, and complicated affair, and it can be managed with models.

In the *Republic*, Plato demonstrates an awareness of this fact when he models rational agency using the structure of a city. His aim was to determine the nature of a just soul, and his method was to “first find out what sort of thing justice is in a city and afterwards look for it in the individual, observing the ways in which

the smaller is similar to the larger” (Plato, 1997b, *Republic*, 369a).¹ What Plato noticed was that agents are complex in a way similar to the way in which cities are complex and, consequently, the former can be modeled using the structure of the latter. Just as the parts of a city—for example, its rulers, merchants, police, warriors, and academics—must work together to produce a city that acts justly, so too, the parts of an agent—for example, desires, goals, emotions, and so on—must work together to produce a just individual. Plato’s idea, then, was to use a representation of the unity required for a just city in order to better understand the type of unity needed to produce similarly just agents.

But Plato may not have realized, or adequately appreciated, that modeling one thing with another introduces distortions. Indeed, in each of the models I have mentioned so far, certain aspects of the target (the thing being modeled) become distorted. For example, the DNA model developed by Watson and Crick distorts the structure of DNA by representing its individual chemical components using cardboard cutouts. Similarly, a physiological model designed to show the location and arrangement of blood vessels distorts its target by using plastic parts. Such distortions, however, are typically not problematic, since we do not expect models to be identical to their targets. Nevertheless, if we fail to acknowledge the differences between a model and its target, and ignore the distortions that arise in the model due to those differences, we may confuse distortions for features of the target. The point is especially germane to Plato’s urban model of agency. While it seems obvious that there are clear differences between cities and individuals—for example, individuals are not composed of subagents in the way that a city is composed of particular individuals—unless we acknowledge those differences, we may confuse distortions of the model for aspects of the target, ending up, rather implausibly, with views of agency that appeal to homunculi.

Now, if it is true that modeling introduces distortions as an effect of representation, we should expect that fully representing the various features of a complex target will

¹References to the works of Plato are by dialogue title and standard Stephanus number. The date in the main text refers to John M. Cooper’s *Plato: Complete Works* (1997), where the translations used here are to be found.

require a variety of models, each representing some features and distorting others. This follows from the fact that no particular model can represent every aspect of its target in every detail. If that is right, we should guard against the idea that a complex system like rational agency can be represented with one, overarching model. Just as it would be a mistake to think that a single model of human physiology can represent it in every detail, or that a particular model of the DNA molecule can represent all of its complexity, so too, it would be a mistake to think that a single model of agency—for example, Plato’s urban model of justice—could represent it in every detail.

If we recognize that models have limitations and distortions built into them, why use them to investigate rational agency? And what implications does a modeling approach to agency have for a broader set of philosophical interests? In what remains of this introductory chapter, I want to address these questions. Before I do, however, we need to get clear on some terminology and also consider the different aspects of rational agency that may need to be modeled.

1.2 Models and Theories

There are a variety of strategies for distinguishing models from theories. Which is the right strategy and whether the distinction can be maintained once made are issues that have generated a bulky body of literature. I am going to resist engaging that literature and, instead, adopt a view that emphasizes the role of ‘theory’ and ‘model’ in scientific practice.² Historically, philosophers of science have emphasized

²My way of speaking does not fully comport with either the received (syntactic) view of theories or its more recent counterpart, the semantic view. The former is a historical relic that was a central component of logical positivism. It is distinguished by its commitment to the idea that statements of theory govern scientific thought and that theories are deductively closed sets of sentences. A consequence of this way of thinking about theories and their role in scientific thought is that models are viewed as subsidiary components of science, fully derivable from the content of specific theories and not required for scientific investigation. The syntactic view was heavily criticized during the latter half of the twentieth century, and contemporary theorists have all but dismissed it in favor of one or another version of the semantic view, which is less a single view than a family of views. Two notable proponents of the received view were Rudolf Carnap and Carl Hempel. For their views, see especially Carnap (1939) and Hempel (1965). Suppes (1977) provides a collection of essays detailing the history and eventual displacement of the received view.

In contrast to the received view, the family of semantic views share the idea that the received view should be turned on its head. Rather than positioning theories at the center of scientific thought, semantic views argue that models (and model building) govern how we think about scientific problems. Influential theorists who have developed such views include Patrick Suppes (1960) and

the relation between theory and model, and endorsed the idea that “good theory already contains all the resources necessary for representation [or model construction]” (Cartwright, 1999, 245). Confidence in that view has recently waned, however. Margaret Morrison, for example, has argued that emphasizing the connection between theory and model is a mistake.³ In coming to this conclusion, she seems to have been influenced by Nancy Cartwright’s arguments to the effect that “the corrections needed to turn the models that are provided by theory into models that can fairly accurately represent phenomena in the physical world are seldom, if ever, consistent with theory” (1999, 251).⁴ For Cartwright, theories lie, and must be manipulated to produce usable models of local behavior.

Nevertheless, even for Cartwright, models and theories seem to be, in some sense, connected. As a result, for her, the term ‘theory’, as with semantic and syntactic views, ends up meaning something like “the general background ideas that influence the construction of models.” I want to retain Cartwright’s idea that theories are general and that they guide the construction of models, while remaining neutral about the role and priority of ‘theory’ and ‘model’ in scientific practice. It is familiar, for example, to think of plate tectonic theory as the background idea used by geologists to build individual geological models. Nevertheless, even if theory and model often

Bas van Fraassen (1980). According to them, models are designed to address different scientific problems, and theories act as background ideas connecting models with the systems or objects they aim to represent. As such, the notions of ‘theory’ and ‘model’ remain tightly connected.

Initially, much effort was spent trying to explain how the relation between models and theories was to be understood, but as these views have evolved, that task has faded from focus. This is in contrast to earlier accounts. Whereas early proponents of semantic views targeted the older, syntactic view by proposing alternatives focused on understanding the notion of theory, more recent views have shifted that focus in an effort to understand and explain scientific thought. A nice example of this transition can be found in Ron Giere’s *Explaining Science*. There he proposes a semantic view of theories that is comprised of two elements: “1) a population of models, and 2) various hypotheses linking those models with systems in the real world.” After proposing the view, however, he immediately acknowledges that for him, “scientific theory turns out not to be a well-defined entity” (1988, respectively 85, 86). Most recently, semantic views amount to little more than the claim that “model construction is an important part of scientific theorizing” (Downes, 1992).

³In “Models as Autonomous Agents” (1999, see especially, 43), Morrison argues that models are functionally independent of theories in the production of scientific knowledge.

⁴The quote is Cartwright’s own characterization of the argument she put forward in *How the Laws of Physics Lie* (1983).

travel together, they need not: we can build models without relying on theories, and we can have theories with no models.

So for my purposes, what is the difference? Following Mark Wilson’s⁵ suggestion that “Mild-mannered ‘theory,’ in its vernacular and scientific employments, often connotes little more than ‘an intriguing proposal’” (2006, 127), I want to suggest that just about any interesting hypothesis can serve as a theory, if it is taken to be true. Nevertheless, in the sense I want to adopt, theories are intended to provide principles that govern objects. As it relates to the topic of this chapter, then, any interesting view about rational agency can be understood as a theory, if it is taken to provide a true statement about the principles that govern such agents. For example, if we assume that instrumentalism—the view that all reasons for action are means-end reasons—is a theory of rational agency, the principle of action governing such agents will be one wherein agents act, and should do so, only in an effort to satisfy or move closer to fulfilling specific desires.

In contrast to theories in this sense, models aim to represent only specific aspects of their targets, and, as a result, distort others to one degree or another. One consequence of the difference between ‘theories’ and ‘models’ in the sense I am using them is that true theories remain true when conjoined, but models, because they accurately depict certain aspects of their targets while distorting others, are not amenable to conjunction. The difference can be stated as follows: true theories conjoin to form consistent, unified wholes; accurate models, because they contain distortions, do not. It follows that modeling all the features of a particular target will require a number of different models.

1.3 Elements of Agency

Before proceeding, it is worth considering in slightly more detail what it is we are trying to model. To be an agent is to be the kind of thing that can produce effects in the world—that is, it is to possess a capacity for self-movement. Understood in

⁵The sense of ‘models’ I am endorsing is closely allied with Mark Wilson’s notion of ‘facades’. For Wilson, facades resemble the “overlapping and fibered sets of maps included in an atlas” (2006, 293). The idea is that just as an atlas requires a variety of representational facades to adequately cover the interests of map users, so too, adequately representing any complex system requires disjoint, but still connected, patches of representation.

this minimal sense, there are many kinds of agents: insects, fish, and primates are agents, but less obvious candidates include businesses, corporations, and government institutions. Of course, human beings are agents too, but our agency has a peculiar feature: its source can be traced to principles of rationality. That is, we are not agents merely: we are rational agents whose movements can be initiated and governed by reasons. As a result, we should look at the potential sources of reasons to get a better sense of what we hope to model.

1.3.1 Instrumentalism

Agents are often moved to act in virtue of ends that they desire to achieve. In fact, one need not look too far afield to see that many kinds of agents tend to be motivated by their aims: spiders move in order to catch prey, apes beat their chest in order to ward off competitors, and human beings drink coffee in order to stay awake. In acting purposefully, even the simplest of agents do things in order to realize their aims, wants, or goals.

Despite this fact, however, there seems to be a difference between the kind of action exhibited by a spider and the kind of deliberative action characteristic of an individual attempting to, say, realize his life's ambitions. Yet impartially articulating that difference has been notoriously difficult. As Harry Frankfurt warns:

While the general conditions of agency are unclear, it may well be that the satisfaction of these conditions by human beings depends upon the occurrence of events or states which do not occur in the histories of other creatures. But we must be careful that the ways in which we construe agency and define its nature do not conceal a parochial bias, which causes us to neglect the extent to which the concept of human action is no more than a special case of another concept whose range is much wider. (Frankfurt, 1988, 78-79)

Even with this precautionary note, however, Frankfurt has argued that there is something peculiar about the type of agency characteristic of persons. Indeed, that idea has received significant uptake, and is now a familiar feature of hierarchical models of rational agency. According to such models, human agents, as persons, are unusual in that they are able to reflect directly on the desires, values, and reasons that motivate them, and to change those desires, values, and reasons in an effort to

align them with deeper, more stable aspects of their agency.⁶ Another way to frame that point, which is more germane to the subject of *rational* agency,⁷ is to say that human agents seem unique in their capacity to reflect on their individual motives and to change them in the face of what seem to be deeper features of who they are as individuals. On this way of thinking, what makes human agents rational is their capacity to figure out who they really are, and to do things in an effort to satisfy their deeper sense of self.

Such views of agency demand a fairly robust form of self-knowledge, because they require that individuals acting for reasons be capable of explaining and justifying the things they do by appealing to deep features of their own motivational psychology. Put differently, for an individual like you or I to account for the things we do for reasons, we must know some pretty deep things about our desires and how they serve to satisfy or frustrate our own sense of who we are. It is by reflecting on the motivational attitudes crucial to our own sense of self, then, that sovereign agents locate a standpoint for rational action. But, how, exactly, does knowing information about one's own deep motivational attitudes *justify* particular actions? After all, if self-reflection reveals a vicious agent, presumably that agent is still not justified in acting viciously. To overcome that potential difficulty, we might hope to show that reflection on one's nature as a rational creature inevitably leads one to discover motivational components that are nonvicious and shared by all rational creatures. If knowing something about oneself as a rational agent reveals something about

⁶There are a variety of problems associated with both the reflective standpoint presupposed by the capacity to reflect on lower-order desires and what the relation between that standpoint and lower-order desires must be like to play a significant role in rational agency. In large part, the controversies grow out of Frankfurt's influential article "Freedom of the Will and the Concept of a Person" (1971). In that article, Frankfurt argues that a person is distinguished by higher-order attitudes about first order desires.

Since his initial paper, a number of writers have weighed in on the debate. An up to date defense of a hierarchical conception of agency can be found in Michael Bratman's *Structures of Agency* (2007b). And a notable exception to the idea that persons have hierarchically structured attitudes can be found in Gary Watson's "Free Agency" (2004).

⁷I do not want to confuse two different foci of work on agency. One has to do with our capacity as *free* agents, the other with our capacity as *rational* agents. Whatever the difference between these two different focal points turns out to be, it is the rational variety that is of interest here, and the emphasis in the text is meant to stress that point.

justified action, reflecting on one's motives may serve to provide a peculiarly rational standpoint for action.

Although such a strategy might sound promising, it is worth noticing the lack of consensus about what the heart of our motivational psychology must amount to in order to produce rational action. The values, desires, and preferences that are supposed to frame the deeper sense of self characteristic of rational agents are not obvious, and consequently, different theorists think that different features of who we are as rational creatures should serve as the motivational foundation for rational action. Indeed, when we, as individuals, reflect on what moves us to act, and on what we think should move us to act, we often find that our own motivational psychology is not unified—that is, the deeper sense of self that philosophers like Michael Bratman and David Velleman appeal to as the standpoint of our own rational agency might, in the end, result in disordered agents. How, in the face of such disorder, can we determine what to do? And is there a way to accommodate the variability within and between agents in a single, comprehensive model of instrumental agency?

I do not think so. When we reflect on the standpoint of our own agency to acquire the knowledge needed to justify the things we do, it is far from clear whether a single account can make sense of our instrumental reasons for action. And in fact, I want to argue that it cannot. Instrumental agency is a disunified concept as it applies to human agents. Consequently, to make sense of it requires a variety of models that each depict one or another aspect of the deep reasons we have for doing the things we do.

1.3.2 Emotions

Although instrumentalist conceptions of agency pervade the literature, there seems to be more to the capacity to act for reasons than can be accounted for with such accounts. For example, human emotions play an important role in the things we do, but making sense of that role has frequently required ideas that seem at odds with instrumentalist conceptions of agency. It has recently been argued, for instance, that agents with impaired emotional functioning have difficulty responding to social cues, because they fail to perceive certain important aspects of their social setting (Damasio, 1994). That is not to suggest, however, that emotionally impaired individuals

fail to have some set of relevant ends. Rather, it is to suggest that they become, in a certain sense, blind to proper forms of emotional expression. To make sense of that fact, we seem to need ideas that do not fit well with instrumentalist accounts of rational agency.

Nevertheless, philosophers like Candace Vogler have still tried to make sense of emotionally motivated action by appealing to the familiar framework of instrumentalism. She argues, for instance, that since acts of emotional expression unfold in a manner typical of actions done to realize or accomplish specific ends, the reasons behind them must be articulable with standard, means-end descriptive terminology. Such an Anscombian view of action explanation presupposes that individuals understand the nature of the actions being performed when acting from emotion. As a result, when an individual gives expression to emotion through action, that agent must know what he is doing while doing it. But such a model of emotional action is mistaken, because it ignores the role of emotion in framing the thoughts that move individuals to act. This observation is best understood by looking at acts of artistic expression, and the emotionally loaded terminology used in the description of such actions.

Let me fill that in a bit by saying more about what I have in mind. The metaphor of vision (or perception) is particularly apt when explaining the role of emotions in guiding actions that give expression to how we feel—for example, acts of artistic expression. Emotions seem to have some influence on how agents perceive the world, which is similar to the influence of the five senses on the perception of physical objects. Just as we need beliefs derived from our perceptual apparatus to think, reason, and act in the world—that is, to be rational agents—so, too, we need emotional perception to think and act in a manner typified by forms of artistic expression. In fact, the idea that emotions function as a form of perception, which itself is a precursor to rational agency, has been emphasized by Ronald de Sousa. He writes:

the canons of rationality that govern [emotions] are not to be identified with those that govern judgment, or perception, or functional desire. Instead, their existence grounds the very possibility of rationality at those more conventional levels. (de Sousa, 1987, 203)

The suggestion is that emotions are in some sense required for more considered forms

of rational agency. This includes both practical and theoretical modes of rationality. For de Sousa, an agent cannot figure out what to do or what to believe without emotional input. If that is right, there must be a fairly robust connection between emotions and rational agency: nothing that we can rationally do or believe is imperious to the influence of emotion.

de Sousa is not unique in his commitment to the role of emotions in rational agency. Many others working in the area have suggested a similar link between the emotions we feel and the reasons we have for acting.⁸ Indeed, there is a growing cadre of philosophers who think that emotions play an important role in our capacity to figure out what we should, or should not, do. As Martha Nussbaum has pointed out:

the *correct* perception of a practical situation requires emotional . . . activity, [and] the emotions have a valuable informational role to play within the ethical life as forms of recognition. (Nussbaum, 1988, 230, emphasis mine)

But Nussbaum may be understating things here. After all, it is clear that emotions influence expressive actions—for example, forms of artistic expression—and that the capacity to understand and express one’s emotions in an appropriate manner deeply influences the shape of an individual’s life.

Nevertheless, it is not clear that we really understand the influence of emotion on the things we do. If it is true that emotions lie behind, and give shape to, the expression of our capacities as rational agents, it may not be possible to fully grasp what we are doing when we are acting from emotion. Indeed, I will suggest that because emotions shape how we see the world, their influence on what we do may be all but invisible to us as intelligent agents. In fact, I want to argue that agents who act from emotion do so without knowledge of what they are doing. This is in sharp contrast to the type of Anscombian model endorsed by Vogler. When we act to give expression to emotion, we are frequently coming to terms with unfamiliar modes of thought, which are tied to the nature of emotions themselves, and require

⁸There is an ever-growing body of literature aimed at understanding the role of emotions in the life of rational agents. A widely used and highly regarded starting place can be found by looking at Amelie Rorty’s *Explaining Emotions* (1980). More recently, Peter Goldie’s collection, *The Oxford Handbook of Philosophy of Emotion* (2010), deals with a number of philosophical issues related to the emotions.

doing things without practical knowledge. If that is right, we should expect such actions to require a different model of action explanation, one that makes sense of the perceptual shifts characteristic of emotional experiences without requiring the type of Anscombian knowledge built into Vogler’s view. Such actions stand as a class of counterexamples to the idea that rational agents must act with practical knowledge, and provide reason for Anscombian like Vogler to reconsider their commitments.

1.3.3 Practical Learning

A further feature of agency that needs to be considered when providing an account of human agents as rational agents is the capacity to discover, learn, and develop into the type of intelligent creatures that we seem to be. After all, an individual’s capacity to be motivated by and act for reasons seems to be something that develops over time. Yet explaining how we develop the capacity for rational action is a problem that does not have an adequate philosophical explanation.

The problem of how we develop new capacities is, I think, analogous to one found in more theoretical realms of philosophy. In particular, it is related to the familiar problem used to motivate Plato’s *Meno*. The dialogue begins with Meno asking, “Can you tell me, Socrates, can virtue be taught?” (Plato, 1997a, *Meno*, 70a). Eventually, Socrates gets around to stating the full force of the problem as he sees it:

[A man] cannot search for what he knows—since he knows it, there is no need to search—nor for what he does not know, for he does not know what to look for. (Plato, 1997a, *Meno*, 80e)

The puzzle presents a challenge for anyone trying to account for the capacity to learn new truths. How can we find what is true, if we do not know what we are looking for? Although we need not worry about this more theoretical version of the problem, or the Socratic solution to it, we should notice that there is a related problem in the practical realm. How do agents learn to perform actions they have never before performed? This question is about the relation between what we already know how to do and what is required to learn how to do new things—that is, how to develop our capacities as agents. It makes sense, then, to look for answers to that question by looking at work that explores that relation. Two areas of research suggest themselves: one is tied to Gilbert Ryle’s distinction between knowledge-that and knowledge-how; the

other is the body of literature pertaining to practical knowledge, which has recently received considerable attention from followers of Elizabeth Anscombe. Both areas approach problems pertaining to an individual's capacity to learn how to perform new actions by emphasizing the role of knowledge in governing the things we do. For example, writing in response to Ryle, Jason Stanley has recently suggested that:

knowing how to do something is the same as knowing a fact. It follows that learning how to do something is learning a fact. For example, when you learned how to swim, what happened is that you learned some facts about swimming. (2011a, vii)

On Stanley's view, facts, or true propositions, are what is known by an agent when he or she knows how to do something. If that is right, knowing how to swim is knowing the propositions relevant to swimming, and learning to swim is similarly a matter of learning the relevant propositions.

The role of knowledge in learning new skills can also be found in the work of recent followers of Anscombe. Sebastian Rödl, for example, has argued for a very strong connection between thought and action. Indeed, he believes that thought and action are so securely connected that thinking about what to do may result in the power to do it. Since knowing how to do something amounts to knowing the information pertaining to doing it, and since information about how to do things can produce knowledge-how, learning how to do new things may simply be a matter of drawing a particular type of inference.

Such views seem to me to be mistaken because they depend too heavily on the role of practical knowledge in action. The connection between what we know how to do, how we learn to do new things, and our capacity to act as rational agents is one that needs to be explored. In fact, I want to argue that the things we do to acquire knowledge-how depend on a class of actions that have previously been ignored. What I call 'explorative actions' are actions that agents perform without knowing what they are doing. But such actions are not mere thrashing; they are governed by a practical analog of induction, and they should be modeled in that way. Such a model promises to provide insight about how we, as rational agents, learn to do things we have never done before, and also promises to provide reasons for reconsidering the Anscombian view that practical knowledge is a precondition of intelligent action.

1.3.4 Autonomy

Finally, to model rational agency, we must account for our sense of autonomy—that is, the sense that we are individuals capable of governing ourselves according to principles of reason. To be autonomous as rational agents, the capacity for self-governance must be tied, not just to reasons, but to reasons that are supported by principles of rationality. Indeed, this is a point made famous by Kant. He writes:

Only a rational being has the power to act in accordance with his idea of laws—that is, in accordance with principles—and only so has he a will. Since reason is required in order to derive actions from laws, the will is nothing but practical reason. (Kant, 1997, 80)

Or, a bit later:

The will is conceived as a power of determining oneself to action in accordance with the idea of certain laws. And such a power can be found only in rational beings. (1997, 95)

In these passages, Kant expresses the idea that a rational agent's capacity to act depends on recognizing law-like grounds for acting. In fact, this is the idea behind Kant's two-pronged universalizability test. Unless an agent's maxims are both conceptually and volitionally consistent, it is not rational for the agent to act on them.⁹ Another, more familiar, way of putting this point is to say that the principles of action governing rational agents must be categorical.

Christine Korsgaard has taken up these Kantian points, putting them to work in understanding problems of agency. She writes:

the principle of governing oneself by universal laws is the constitutive principle of rational activity generally. For the requirement of universalizability governs every aspect of rational thought. To believe on the basis of a rational consideration is to believe on the basis of a consideration that could govern the beliefs of any rational believer, and still be a belief about the public, shared world. To act on the basis of a rational consideration is to act on the basis of a consideration that could govern the choices of any rational chooser, and still be efficacious in the public, shared world. (Korsgaard, 2008, 12)

For Korsgaard and Kant, our autonomy as rational agents—that is, our capacity to rationally govern ourselves—is grounded in particular principles of rationality.

⁹One of the clearest discussions of Kant's universalizability test, and its role in determining the moral value of an action, can be found in Onora O'Neill's "Universality Tests in Autonomous and in Heteronomous Ethics" (2001).

Only agents who act on the basis of considerations that are universalizable qualify as autonomous beings.

But the idea that rational agents must be moved by maxims that meet certain strictures of rationality raises a problem for an approach to agency that, like mine, relies on models. As I have already noted, models have limitations and distortions built into them. It is unlikely, then, that it will be possible to conjoin several models to produce a single, coherent account of rational agency. In other words, given the fact that the scope of models is limited, and that they distort their target in one way or another, it is improbable that they will be able to be brought together to form a single, coherent picture of rational agency. This is problematic if we aim to make sense of autonomy, and rational agency more broadly, in terms of principles that are universalizable. For, if nothing else, universalizable principles must be coherent and unifiable under a single, overarching principle.

The concluding chapter will, to a small extent, address some of these worries. For example, what hope is there for creatures like you and I to unify our agency around principles of reason? And what are the prospects for integrating the models of agency discussed in this dissertation? If they are not significant, as I will suggest, what should we say about the control we have over the things we do as intelligent creatures who act for reasons?

1.4 Is Agency Worth Modeling?

With a better sense of what it is we are aiming to make sense of using models, we can now consider whether a clearer picture of rational agency will be worth the effort. After all, if we are going to expend the energy, should we not expect some kind of payoff? To answer that question, I want to look at what implications modeling agency might have for related areas of philosophy.

1.4.1 Practical Reasoning

Since representations of rational agency purport to model features of *rationality* that move us to act, they will, in one way or another, have implications for how agents should think about what to do. This feature of the project brings it into close contact with theories of practical reasoning. As a philosophical discipline, practical reasoning

is about figuring out the patterns of inference legitimate for determining how one should act. Presumably, these patterns are canons of rationality, which means that reasoning done right is reasoning done according to these patterns.

Be that as it may, we need more information about human agents as rational agents to determine how, exactly, theories of practical reason should be applied to human beings. After all, to show that a pattern of inference is valid is not yet to show that human agents can or should use it to figure out what to do. For instance, if the patterns of inference necessary for realizing long-term plans are too complex to use, it would hardly make sense to insist that they be used to determine what to do. Similarly, if the joint application of incompatible patterns of inference cannot be ruled out, we may not be justified in using either one when figuring out what to do. If these speculative comments are right, then recognizing that a pattern of inference is rational is not sufficient to show that we should use it.

If we acknowledge that specifying a form of inference is not sufficient to show that it should be used, we need to figure out a way to determine which inferential patterns we should use when thinking about what to do. To my mind, a good way to proceed with that task is to get a better sense of the type of creatures we are, the type of agency we possess, and the limitations we face as rational agents. By better understanding ourselves as creatures who act for reasons, we might gain a better sense of what forms of practical inference should or should not be implemented by distinctively human agents. One way to do that is to design a variety of models to capture the seeming variation amongst human agents.

1.4.2 Ethics

In contrast to practical reasoning, which is the project of determining the patterns of inference needed for thinking about what to do, ethics is about what agents should do. More colloquially, practical reasoning is about how individuals should think, ethics is about how individuals should live. Put in those terms, there seems to be a natural link between the two: if living a good life requires thinking about what to do, figuring out the patterns of inference useful for determining what one should do seems important. Consequently, to the extent that models of agency will inform practical reason, it will also improve our thinking about peculiarly ethical problems.

Aside from that connection, however, the idea that we should use models to better understand agency has more direct implications for the field of ethics. For some time, ethicists have been faced with a division between two unyielding camps. On the one hand, there are those who insist on the idea that rational agents should act only with an eye toward the likely consequences of their actions. Accordingly, only if an action is likely to produce the best potential outcome is one justified in performing it. On the other hand, there are those who think that rational agents should act only on intentions that are properly formed. According to this line of thought, only if an agent's intentions are determined by principles of rationality is an individual justified in acting from them. But notice that these two approaches to ethics rely on different ideas about the nature of rational agency. Indeed, they rely on peculiarly narrow conceptions of the nature of rational agency. To the extent that we can gain a better understanding of ourselves as rational creatures using a variety of models, we will be better positioned to engage a range of ethical debates, and maybe shake up traditional approaches to ethical questions.

1.4.3 Social and Political Philosophy

Modeling agency also has implications for social and political philosophy, which are areas of philosophy largely concerned with the social institutions and practices that influence the shape and quality of human life. Indeed, as Plato noticed early on, the design, structure, and function of social institutions—for example, the governing body of a city—are similar to the design, structure, and function of human agents: both are constituted of competing parts that must work together as a unit. Understanding human agency, then, might help us to gain a better understanding of social institutions.

In fact, the agent-like features of corporations, government institutions, and other social and political actors make this area of philosophy dependent on a robust understanding of rational agency. If these large institutions are in fact agents who do things for reasons, then understanding what moves them to act will depend on understanding the forms of rational agency relevant to them. Of course, like their human counterparts, social and political institutions are not simple systems.

They too are complex. And as I have been suggesting, when dealing with complex systems, modeling proves beneficial. For these reasons, then, the kind of approach I am proposing should be useful for social and political philosophy.

1.5 Why the Modeling Approach?

Granted that a better understanding of agency has implications for a variety of philosophical projects, why use models and modeling as the method of investigation? If models represent only certain aspects of a target, inevitably contain distortions, and are each insufficient to provide a complete picture of agency, why bother? Should we not prefer a strategy that promises a comprehensive and undistorted picture of our nature as rational creatures?

Possibly, but a method sold with that promise may be over-selling itself. Consider the nature of representation itself. As Plato suggests in the *Republic*, and as I pointed out earlier on, representations have limits. Indeed, Plato argues that a single painting of a bed cannot truly depict the features of an actual bed, because a representation cannot represent aspects of the thing it is supposed to depict (Plato, 1997b, *Republic*, especially 598c). As that point applies to our topic, we should be wary of a strategy that promises a single, comprehensive, and undistorted picture of rational agency. That point is trivial if all I mean by it is that we should be wary of written accounts of rational agency because they cannot, for example, truly depict the movement of agents without themselves being in motion. But that is not what I mean. My point, rather, is that no matter how we choose to represent ourselves as rational agents, we will inevitably be forced to leave things out that are necessary for a complete picture.

Consider, for example, the fact that we are self-conscious agents: we are aware of the fact that we are creatures who know how to move ourselves. Suppose I want to represent that feature of agency, and so build a model to do so. Well, in virtue of focusing on that feature of agency, I may have to give up representing other features of agency that are not compatible with it. For example, I may have to abandon the hope of representing the influence of the subconscious, the influence of emotions, or the influence of factors external to an agent's psyche on the reasons that move individuals to act. But if these alternative influences play an active role in shaping

agency, designing a model to represent our self-consciousness may end up being only a partial representation. In fact, it seems that frequently what we know about ourselves, and our nature as rational creatures, is only part of the story. Consequently, we should be cautious of approaches to agency that promise a comprehensive and undistorted picture.

Let me try to make that point a bit differently while sticking to the example from Plato. The Platonic form of being a bed and paintings of beds seems to bear something of an analogous relation to the idea of rational agency and representations of rational agency. Just as a representation of a bed cannot capture aspects of the Platonic form of being a bed, so too, representations of agents as rational cannot capture certain aspects of what is required to be a rational agent. The difficulty stems, I think, from the fact that human agents must sometimes act without fully understanding what they are doing. There are aspects of our nature as rational agents that are, to one degree or another, hidden from our own sense of what it takes to be rational. Yet they are required for us to be the type of rational creatures we presume to be. As a result, standard representations of rational agency tend to leave these features out.

So how do we overcome the difficulty? One way is to use a multitude of representations to capture all the features of the thing being represented. In the case of Plato's bed, this means getting a better idea of actual beds by, for example, representing them from a variety of angles using a variety of artistic media in an effort to portray a multitude of features. This, of course, runs parallel to the strategy I propose to use in this dissertation. But the alternative I am criticizing promises something different: it promises to bring representations of rational agency together to form a comprehensive, undistorted account of what it means to be a rational agent. This seems to me to be an implausible alternative. After all, what principle of unification could bring a variety of *representations* of a bed together to form a unified whole? And, similarly, what principle of unification could bring a variety of *representations* of agency together to form a unified whole?

These potential difficulties, however, are not the only reasons to be cautious of views that promise comprehensive and undistorted representations of rational agency.

In addition to the fact that such strategies are unlikely to succeed, they also tend to stifle philosophical innovation. There are at least two reasons for believing this. First, features of rational agency that are accepted as true tend to lead to the rejection of accounts that are incompatible with those presumed truths. So, for example, if rational agents *must* have consistent beliefs, it would seem absurd to accept an account that is incompatible with such a presumed truth. Consequently, any potentially innovative model at odds with it would tend to be dismissed out of hand. If rational agents must have consistent belief sets, and a model is proposed that does not meet this condition, why take it seriously? If one is committed to arriving at a comprehensive and undistorted representation of rational agency, the answer to that question is that one should not. But by pursuing a model-based strategy, we can avoid this problem, since such a strategy requires individuals to recognize that their own models cannot be entirely correct.

Another reason to think that approaches that are not model-based will stifle innovation is because they tend to be top-down. Too often, philosophy begins with a fairly simple idea and attempts to explain as much as possible given that idea. But such a strategy often ignores nuance for the sake of simplicity, and tries to deal with complex issues by fitting them to precast philosophical molds. This is particularly evident, as I have suggested, in ethical debates. A stubborn commitment to either consequentialism or deontological perspectives frequently drives philosophical thought. Rather than letting counterexamples to either perspective drive more nuanced views, we try to fit the counterexamples to the top-down perspective. But since we are nuanced agents, representing the sophistication of ethical creatures will likely be handled better by a method that is suited to deal with complexity. By designing models that represent particular features of rational agency, we let the problems drive our thinking rather than our preconceived thoughts drive the problems. If we are forced to think about what our models are for, what features of a target we want to represent, what distortions will result from our representations, and how they might be improved, our philosophical views promise to be more rich, innovative, and useful than their top-down alternatives.

1.6 Prospectus

It is frequently assumed that rational agents are unified agents. As a normative feature of agency, the idea suggests that understanding the things we do requires understanding our reasons for doing them as reasons that can be represented with a single, comprehensive view. But I do not think that any such comprehensive view can be provided, and I want to show why by arguing that features of our own agency undermine attempts to do so. To be the type of creatures we are, we have to act in ways that cannot be captured by a single account of rational agency. Indeed, making sense of ourselves as rational agents requires a number of different models that cannot be nicely fitted together. If I am right, the upshot is that a unified representation of rational agency may not be possible.

In the next chapter, *Instrumental Facades*, I focus on a collection of views that require agents to act with practical knowledge—that is, with an awareness of what they are doing when doing it. I argue against three proposed theories of instrumental agency by arguing that they are in fact better interpreted as models that represent importantly different aspects of rational agency. After sketching the views of Michael Smith, David Velleman, and Michael Bratman, I argue that the heart of each view is needed to represent human agents as instrumentally rational. We cannot give these views up without giving up resources needed to explain how we deal with practical issues we all have an interest in managing. Yet each view is normatively incompatible with the others. This leads to a puzzle: how should we represent nonoptional features of human agency that prescribe actions that are at odds? In answer to that question, I argue for a model-based perspective for thinking about instrumental agency, and provide support for the idea that it is a disunified concept.

In contrast to instrumentalist views that require agents to act with practical knowledge, *Modeling Expressive Actions* aims to develop an account of rational agency that does not require knowing what one is doing when acting. This third chapter is meant to challenge the Anscombian view that intentional actions presuppose practical knowledge. The challenge depends on locating a class of actions that agents purposefully perform without such knowledge. I use acts of artistic expression as an example of agents who act without knowing what they are doing, and argue that many of

the things individuals do from emotion are similar to the actions of artists. I call these actions ‘expressive actions’ and show that they tend to resist being described in a manner that demonstrates an awareness of what is being done while doing it. Indeed, I suggest that giving expression to emotion is an act of clarifying and ordering what one feels through action, and that the clarity that comes through such action cannot be explained using Anscombian views like the one developed by, for example, Candace Vogler.

In Chapter 4, *Acquiring Knowledge-How*, I take the idea that we sometimes must act without practical knowledge a step further by arguing that explaining the capacity of rational agents to acquire practical know-how—that is, for agents to learn how to do things they have never done before—we need another model of agency built around the idea that we perform intelligent actions without knowing what we are doing. Again, this presents a new challenge to the Anscombian view that the only way to act intelligently is to act with practical knowledge. Indeed, I argue instead that much of what we do when learning how to do new things requires acting without that type of knowledge. Only in this way can we develop the skills we need to be the rational creatures we are.

In the short concluding chapter, I develop the implications of the previous chapters for the idea that we must, in some sense, be unified agents in order to act intelligently. To make sense of a variety of features of our own agency, we need distinct models that cannot be nicely conjoined using a broader framework of representation. I suggest that this fact is an indication that we are both unified and disunified in a peculiar sense, and conclude with an argument to that effect. If I am right, we have some reason to be skeptical of views that require that rational agents act according to self-determined laws of rationality.

CHAPTER 2

INSTRUMENTAL FACADES

2.1 Introduction

What are the grounds of rational agency? Instrumentalists answer that question by appealing to pro-attitudes and beliefs, which together are thought to cause, explain, and rationalize the things we do for reasons (Davidson, 1980, 3-4). Thus, when an individual acts for a reason, she intentionally¹ *x*'s in order to *y*, where an inclination towards realizing *y*, along with the belief that *x*-ing is a means to achieving *y*, causes, rationalizes, and explains her action. But that is only half the story. In addition to describing the grounds of action, accounts of rational agency are meant to be normative—that is, they aim to show what it means to be right, correct, justified, or rational in acting. A view of rational agency, then, has two aspects: one explains what rational agents in fact do by describing their reasons, the other explains what rational agents should do by prescribing standards of correctness.

Attempting to square these two aspects of rational agency in order to account for the reasons that do and should move us to act is difficult. Indeed, describing the central attitudes and capacities that move creatures like you and I to act tends to produce a variety of views of agency that conflict in a range of cases. Yet many of these attitudes and capacities seem to be nonoptional—that is, we need a variety of them to make sense of problems that we all have a stake in managing. The result is a collection of views that rely on seemingly nonoptional features of agency and provide conflicting normative guidance. If that is right, how should we think about human agents as instrumentally rational? I am going to argue that we should think in terms of models.

¹The concept of intentionality carries with it significant philosophical baggage. Despite this fact, however, I want to simply assume for now that when an individual acts in order to accomplish an end, goal, or aim, he acts intentionally—that is, with an awareness of what he is doing when doing it. I am going to call that awareness in action ‘practical knowledge’ and, as we proceed through this chapter, I will point out where it crops up in the views to be surveyed.

To understand human agents as instrumentally rational requires a variety of different representations, which are needed to show the connection between what we in fact do and what we should do. These different models, or facades of representation, cannot be integrated into a single, unified view. The upshot is that a complete account of instrumental agency, one that explains human agents as instrumentally rational, will be disunified.

To make the argument, I am going to use three different views of instrumental agency to show that different intellectual capacities and pro-attitudes are needed to describe human agents as instrumentally rational. Michael Smith, Michael Bratman, and David Velleman each provide views of rational agency that rely on capacities and attitudes that are important for understanding our own instrumental rationality. I spell out what this means and defend my reasons for believing it in Section 2.2 of this chapter. In Sections 2.3 and 2.4, I argue that the views of Smith, Bratman, and Velleman provide conflicting normative guidance in a range of cases, and that, nevertheless, they are each nonoptional in the sense that, if instrumentalist views are going to have any chance of fully explaining the reasons that do and should move human agents to act, the features built into these views need to be accommodated.² This leads to a puzzle: how do we square the fact that nonoptional features of our own instrumental agency lead to normative theories that conflict on particular occasions? I argue that the puzzle cannot be resolved by integrating these views into a single view of instrumental agency. Instead, responding to it means recognizing our own complexity as agents and the disjoint facades needed to represent that complexity. I use insights from the philosophy of science to support that idea. As it happens, there is often hidden complexity in the things we aim to understand, which is best represented using a patchwork of facades.

²Notice that I am not suggesting that these specific accounts are themselves *necessary* for understanding rational agency. Nor am I endorsing the broad contours of instrumentalist views of rational agency. Rather, I am simply pointing out that, to the extent that instrumentalist views of agency successfully account for human agents as rational agents, the features at the heart of each of the views I survey in this chapter will be needed.

2.2 Three Views of Instrumental Agency

For each of the following views, their descriptive and prescriptive aspects depend on intellectual capacities and psychological attitudes that are nonoptional. In this section, I describe the capacities and attitudes at the heart of each view. In the sections that follow, I argue that the accounts provide incompatible normative guidance, and that the psychological attitudes around which they are built cannot be abandoned—that is, they are nonoptional. If I am right, the heart of each view is indispensable for a complete picture of human agents as instrumentally rational, and each view is also in normative conflict with the others.

Standard instrumentalism is a view centrally committed to the means-end relation and tied to an agent's psychology via beliefs and desires. The latter (desires) provide agents with motivation; the former (beliefs) show what agents must do in order to get what they want. Such a minimalistic picture of rational agency might lead one to wonder: where is the rationality? Bernard Williams provided one kind of answer to that question. He argued that reasons that explain action must be tied to an agent's subjective motivational set—that is, to the set of things an individual could find worth doing by reflecting on what she already finds desirable (Williams, 1981a, 102). On his view, an agent's rationality is evident when she reasons cogently from her particular set of motives to determine what she has reason to do (1981a, 102–103). The result, as Michael Smith puts it, is that Williams's

conception of reasons, like Hume's own, is predicated on a scepticism about the scope for reasoned change in our desires; predicated on denying that, through a process of rational deliberation . . . we could ever come to discover reasons we all share. (Smith, 1994, 165)

In other words, it is rational to do whatever we want to do as long as we figure out what we want to do in a rational way—that is, as long as we use correct forms of inference and true beliefs to determine what we want to do.

The picture offered by Williams is one in which what an agent is moved to do, and what the agent should do, relies on an individual's beliefs and desires: an agent's nature as a rational creature is described by the psychological states that justify his actions. Anything anyone does for a reason can be described *and* justified by appealing to the states of the individual: if an action is done because of a motive

arrived at via rational processes of deliberation, it is a rational action. This feature of Williams's view has resulted in a good deal of criticism because it fails to explain important aspects of our nature as rational creatures.³ As a result, each of Smith, Velleman, and Bratman has put significant effort into avoiding the sort of difficulties found in Williams's view.⁴

2.2.1 The Rational Ideal

I begin with Smith's view. Contrary to Williams, he argues that the very idea of a normative reason is incompatible with a relativistic perspective. Smith writes:

what my actual desires are to begin with is, on this [Williams's] relative conception of reasons, an entirely *arbitrary* matter, one without any normative significance of its own. [The desirability of a consideration] thus turns out to be an entirely arbitrary fact about it. But arbitrariness is precisely a feature of a consideration that tends to undermine any normative significance it might initially appear to have. (1994, 172-173, emphasis in original)

The point is that if normative reasons depend on an arbitrary set of desires, an agent may have a normative reason to do *anything*. But having a normative reason to do anything is incompatible with the idea that individuals should do some things and not others.

To overcome this problem, Smith argues for a view that grounds normative reasons in beliefs about what fully rational agents would desire. He thinks that by reflecting on the concept of desirability, less-than-fully rational agents will converge on beliefs about what categorical desires are necessary for organizing and living life. Again, from Smith:

The epistemology of value . . . requires the individual to see herself as one among a group of individuals who are trying to answer a common set of

³See, especially, Christine Korsgaard's "Skepticism about Practical Reason" (2001), as well as Elijah Millgram's "Williams' Argument Against External Reasons" (1996).

⁴Each view also depends on a fairly robust notion of what I am calling 'practical knowledge', which is the idea that when acting rationally, individuals must understand what they are doing while they are doing it. The idea that we act with such knowledge will be important as a point of contrast between instrumentalist models and those that I will be developing later in the dissertation. But for now, the issue of knowledge in action can in large measure be placed on the back burner in order to focus on the attitudes and capacities necessary for understanding human agents as instrumentally rational.

questions, questions whose formulation does not require reference to any one of them in particular. ... In deciding which desires to begin from in the attempt to find a systematic justification of our desires, then, we have no choice but to look for desires that are ... widely shared. We have no choice given two key assumptions: first, that the goal is to find a single set of desires that all rational creatures would acknowledge to be systematically justifiable, and second, that none of us has any special epistemic gifts that would justify us in privileging our own desires and judgements over the desires and judgements of others (1994, 176–177)

The method is a version of reflective equilibrium. By considering the widely shared background desires of agents, and denying privileged epistemic status to the desires of particular individuals, Smith thinks that less-than-fully rational agents can figure out what there is normative reason to do.

While we can ignore the difficulties associated with Smith's view about the origins of normative reasons, we should notice that he relies on intellectual capacities associated with pro-attitudes shared by all agents to connect the descriptive and normative aspects of his conception of rational agency. In particular, it is in virtue of shared desires, and a capacity to reason subjunctively, that less-than-fully rational agents are able to determine what fully rational agents would desire. Such subjunctive reasoning results in beliefs that then shape the desires of less-than-fully rational agents, because, Smith believes, an agent's desires are sensitive to beliefs about the desires of their fully rational counterparts. In this way, the desires of human agents can, according to Smith, be used to determine the counterfactual desires of fully rational agents, which set the normative standard for agents like you and I.

But the movement from the desires of less-than-fully rational agents to those of fully rational agents is only possible if we assume shared intellectual capacities and a deep understanding of our own desires. It is because human desires are similar to those of their fully rational counterparts, and because we know what they are and can reason subjunctively about them, that Smith thinks we can derive normative reasons for action from the desires typical of human agents. The claim that “if we believe that we would desire to ϕ if we were fully rational then we rationally should desire to ϕ ” (Smith, 1994, 177) hinges on our capacity to reason subjunctively about desires and beliefs similar to the desires and beliefs of fully rational agents (1994, 165,

for example). And of course, that requires a fairly robust understanding of our own motivational psychology and the reasons that move us to act when acting on purpose.

2.2.2 Autonomous Agents

For Smith to make good on the normative aspect of his view, he will need to say more about why an individual's desires *should* be sensitive to her subjunctive beliefs about the desires of rational agents. Furthermore, even with that issue addressed, Smith owes us a substantive account of normative reasons. He cannot escape that obligation by simply suggesting that "substantive convergence [on normative reasons] is always assumed to be available, in so far as we converse and argue about the reasons we have" (Smith, 1994, 173). If conversation and argument could guarantee convergence on what counts as a normative reason for action, we would expect some convergence after 2,500 years of philosophy. Since we do not see that type of convergence, we should be skeptical that Smith can furnish what he owes.

Velleman seems to agree. He suggests that providing an account of normative reasons from an independent conception of rationality is not likely to be forthcoming (Velleman, 2000, 179). How, then, does Velleman hope to provide an account of rational agency with prescriptive content? Rather than trying to formulate a substantive account of rationality, Velleman aims to locate a feature of agency that can establish the correctness of actions as such. The idea is to determine an object at which all actions must aim in order to be successful as actions. By locating such an object, Velleman thinks he can establish an action's correctness by determining whether it satisfies this aim. Put differently, Velleman's hope is to determine what feature of agency could establish an action as justified by determining what feature of agency all actions must aim to realize. Rather than trying to forge a connection between what agents are motivated to do and what they should be motivated to do by locating features of rationality, Velleman's aim is to bring the descriptive and prescriptive aspects of his view together by locating a feature of agency shared by all agents capable of doing things for reasons.

What, then, does Velleman think the feature of agency is that establishes the correctness of actions? The short answer is that it is the higher-order inclination for self-understanding. As Velleman writes:

As a human being, you are naturally endowed with a theoretical intelligence, which is not a passive receiver of information but an active synthesizer, striving to make incoming information hang together so as to represent an intelligible world. You are also endowed with an objective self-awareness, a concept of yourself as a part of the world to be understood. These two endowments inevitably combine to yield the aim of understanding yourself, which inevitably motivates you to do what you can understand (2009, 136)

This passage is a toned-down version of more controversial themes Velleman explored in *The Possibility of Practical Reason* (2000). There he writes:

Suppose you have an inclination toward being in conscious control of your next move. This inclination will inhibit you from doing anything out of other motives until you've accepted that you're going to—precisely so that you'll do it only after and because you know it, and hence under conscious control. Once you accept that you're going to do something, however, the inclination toward being in conscious control will reinforce your other motives for doing it, since doing what you've accepted you'll do is what puts consciousness in control. Your inclination toward conscious control is thus converted, from an inhibition against doing something into a motive for doing it, by your accepting that you'll do it. (Velleman, 2000, 196)

According to these passages, it is by reflecting on what we believe our desires and preferences to be that we come to have the desires and preferences we have. These desires and preferences, in turn, lead us to act in ways that are intelligible to us. The thought is that by forming beliefs about oneself and one's place in the world, individuals are led to do whatever it is they believe about themselves.⁵ This conception of rational agency, of course, requires a fairly robust commitment to practical knowledge. Indeed, on Velleman's view, it is by acting from an understanding of what one knows about oneself that agents take control of their actions as rational agents.⁶ For example, suppose I believe that I am the kind of person who prefers staying home Friday nights to play World of Warcraft. Reflecting on myself and realizing that this belief represents my preferences, I am led to do what I prefer as I reflect on what I

⁵The terminology being used here might lead the reader to believe that Velleman endorses the idea of a "self" as an essential part of agency. He does not. For Velleman, the self is just "a word used to express reflexivity" (Velleman, 2002, 111).

⁶The role of what I am calling practical knowledge in Velleman's account of rational action is admirably characterized and challenged by Kieran Setiya in "Practical Knowledge" (2008).

believe. So when Friday comes around and I am thinking about what to do, my belief about myself influences what I do, and, sure enough, I find myself comfortably doing what I believe I wanted to do all along. As Velleman puts it:

because I want to do whatever makes sense to me, I can jump to the conclusion that I prefer to do something, since doing it will then make sense to me, and I shall therefore prefer to do it. If my conclusion represents itself accurately, as a self-fulfilling belief adopted out of a desire for its fulfillment, then it will constitute a full-blooded intention. (1989, 181)

For Velleman, what an agent wants to do is determined by deliberating on what she believes about her own desires and preferences.⁷ Once an agent knows what she believes about herself—that is, once she possesses knowledge of her desires and preferences—a higher-order desire to act only on what is intelligible to her engages her beliefs and causes her to act. Only in this way is an action fully in one’s conscious control: what an agent does is done intentionally only when she knows what she desires and is motivated to do what she understands her desires to be. On the view proffered by Velleman, rational agency is agency directed at the aim of self-understanding.

But why *should* an agent act in order to achieve self-understanding? In other words, what is the connection between what we are in fact moved to do and what we should be moved to do on Velleman’s account? We already have the answer: it is the capacity of an agent to be self-intelligible along with the higher-order desire to act intelligibly. An action is done correctly when it is done to realize the aim of acting intelligibly to oneself. Again, from Velleman:

I claim that the criterion of correctness for action is how it would make sense to behave, because action consists in behavior that aims at making sense, by being sensitive to considerations of its own intelligibility, which therefore qualify as reasons for acting. (Velleman, 2009, 135)

An agent has a justified reason to act when what she wants to do is intelligible to her as a reason because of what she knows about herself. Rational human agency, then,

⁷A peculiar result of this move by Velleman is that deliberation about what to do reduces to theoretical deliberation—that is, for Velleman, practical reasoning properly construed is theoretical reasoning. As a result, the mechanism for rationally determining an individual’s ends will be the relation between an individual’s beliefs and the consequences of those beliefs according to valid forms of theoretical inference.

is explained by a particular intellectual capacity associated with a higher-order desire shared by all rational agents.

2.2.3 Planning Agents

Following Velleman, Bratman explains the grounds of rational agency using intellectual capacities and attitudes that are required for intentional action. In contrast to Velleman, however, Bratman thinks these aspects of agency are more varied, complex, and subtle than those argued for by Velleman. Indeed, Bratman emphasizes a trio of features—an agent’s reflectiveness, planfulness, and self-conception as a temporally extended creature—that produce a collection of pro-attitudes at the core of agency, which he uses to argue for an account of rational agency with normative content. In particular, Bratman’s account relies heavily on the fact that self-reflective, temporally extended agents must be capable of cross-temporal organization in order to achieve their long-term goals. The picture relies on a fairly robust notion of self-knowledge: when an individual acts on purpose for reasons, the action is performed from an understanding of the cross-temporal attitudes responsible for holding the agent together. As a result, what agents are in fact motivated to do and what they should be motivated to do are connected by Bratman via the capacities and attitudes required for cross-temporal persistence as an organized agent.

Let me explain. The fact that we are reflective beings allows us to make judgments about ourselves. In particular, it allows us to see ourselves as creatures who are temporally extended, which in turn fosters the idea that we must live with an eye toward the future. Thinking about what to do often is planning for what we will do in the future. To persist through time and to effectively realize one’s future plans, certain psychological attitudes are required. These attitudes, along with the capacity to reflectively organize oneself through time, are required to act for the short, and long-term, future. So, from the fact that agents are reflective, planful, and temporally extended, Bratman argues for a collection of higher-order attitudes that require the capacity for cross-temporal self-organization, which serve to hold the agent together through time.

Of particular concern are the higher-order attitudes that are the basis of an agent’s identity over time. These ‘self-governing policies’ fit together to produce the Lockean

ties that hold individuals together as the same agent. Bratman writes:

[Self-governing policies] support the cross-temporal organization of basic forms of practical thought and action in ways that involve associated Lockean ties of cross-reference and continuity. And this is part of their characteristic functioning. In this characteristic functioning, they help organize the practical life of the agent: they help organize, over time, the agent's practical thinking (including forms of deliberation and planning), the agent's activity, and the complex interrelations between such thought and action; they help constitute and support the temporally extended, interwoven, interlocking structure of coordinated practical thought and action. And they do all this in part by way of constituting and supporting relevant Lockean ties, Lockean ties that help constitute the persistence of that agent as one and the same over time. In functioning in these ways, they help constitute the metaphysical backbone of our temporally extended practical thought and action. And it seems to me reasonable to say that it is in playing these Lockean roles in organizing the temporally extended practical thought and action of one and the same agent, that these attitudes earn the authority to speak for that agent. These attitudes have agential authority *at* a time in virtue of their roles in constituting and supporting the interwoven, interlocking, structures of agency of that person *over* time. (2007c, 245)

According to this picture, the higher-order attitudes responsible for an agent's persistence over time organize thought and action in a manner that speaks for the agent. As a result, when agents think or act according to them, those thoughts and actions proceed from a standpoint that is uniquely the individual's own. In this sense, an agent's higher-order policies justify the things she does.

This is only true, however, if the agent is, in a certain sense, satisfied with her higher-order policies. The idea that higher-order policies justify an agent's actions only if she is satisfied with them provides a path for Bratman to say more about rationality and its role in agency. In particular, since higher-order policies are the attitudes that tie an individual agent together over time, "they differ in basic ways from ordinary desires: in particular, they are subject to distinctive rational norms of consistency, coherence, and stability" (Bratman, 2007a, 27). To hold themselves together over time, agents must resolve potential conflicts between attitudes that define them as the same individual. This, of course, requires knowing what the motivational attitudes that hold one together are. Cross-temporal organization, then, requires self-knowledge. Once these conditions are met, norms of rationality serve to

constrain the attitudes of individuals through time, because “to be satisfied with a self-governing policy ... is for that policy not to be challenged by one’s other self-governing policies (Bratman, 2007a, 44). The picture we have from Bratman, then, looks something like the following: the higher-order policies that tie us together as individuals and organize our thought and action through time require certain demands of rationality to be met. When we are aware of these policies, act in accord with them, and do so with an awareness that we are doing so, we act as rational agents. If that is right, rational agency requires practical knowledge as a temporally persisting agent.

As we see, Bratman’s self-governing policies parallel Velleman’s higher-order desire for control. Just as the higher-order desire for control brings actions done to achieve lower-order desires under the constraints of rational norms, Bratman’s policies provide ends that establish the conditions for acting on what one desires. Because of this, Bratman has a story to tell about when an action is justified. He writes:

To endorse a desire is, roughly, for that desire to be endorsed by attitudes whose role it is to support the temporal organization of agency by way of constituting and supporting Lockean ties characteristic of [one’s] temporal persistence. (Bratman, 2007a, 31)

Now, since the attitudes that support the temporal organization of agency through time are self-governing policies with which the agent is satisfied, Bratman can claim a few pages later that:

An agent’s reflective endorsement of a desire is ensured by a self-governing policy—a policy with which the agent is satisfied—of treating that desire as providing a justifying end in motivationally effective practical reasoning. (Bratman, 2007a, 40)

In other words, an agent is justified in treating a desire as something worth acting on to the extent that it accords with higher-order policies, which meet specific demands of rationality and are responsible for holding the individual together as a temporally persisting agent.

So from the fact that we are temporally extended beings with the capacity to reflect and plan our lives, Bratman derives a conception of rational agency that has normative content. We should intentionally x in order to y when y is endorsed by policies with which we are satisfied. In this way, the things we do for reasons stand in

the service of psychological attitudes required for our own temporal persistence. The descriptive and prescriptive components of Bratman's account, then, depend on an agent's knowledge of the higher-order policies at the heart of agency and a capacity for self-organization.

2.3 Prescriptive Discord

I have argued that Smith, Velleman, and Bratman rely on specific intellectual capacities, and their associated pro-attitudes, to connect what we are moved to do with what we should be moved to do. In each case, it is because agents possess certain capacities, which are tied to the standpoint of their own motivational psychology, that they are able to figure out what they should do and to act with an understanding of what they are doing as rational creatures. To recapitulate: for Smith, the grounds of rational agency are determined by the intellectual capacity of agents to think counterfactually about desires. By determining what desires the fully rational agent *would* have, an individual can figure out what desires she *should* have. Velleman, on the other hand, appeals to the capacity for self-intelligibility and the higher-order desire to understand one's actions. Because intentional action requires the higher-order desire to do what one understands, intentional actions occur only after an individual has made herself intelligible to herself. In contrast, Bratman's more nuanced account is centered around an agent's capacity for cross-temporal organization and higher-order attitudes necessary for temporally extended agency. It is in virtue of the fact that we are reflective, planful, and temporally extended that we adopt policies, values, plans, and intentions to organize our thought and action over time. In each case, when an individual acts, he acts with and from an awareness of what he is doing when doing it.

The intellectual capacities peculiar to each view are straightforwardly distinct. Not only do they each describe substantively different capacities, they also deliver conflicting normative guidance for rational creatures. The result is three discordant portraits of rational agency.⁸ Despite this fact, however, the capacities and attitudes

⁸One natural response to this point is to suggest integrating the discordant views into a single, comprehensive view of instrumental agency. I address this issue in Section 2.5 of this chapter.

at the heart of each view are needed to manage practical issues important to the lives of human agents. I will take up that latter point—that is, the idea that the central attitudes and capacities put forward by Smith, Velleman, and Bratman are nonoptional when portraying the instrumental rationality of human agents—in the next section. For now, I want to provide support for the idea that these substantively different capacities result in discordant normative demands.

2.3.1 Smithian Agents

For Smith, less-than-fully rational agents figure out what they should do by determining what fully rational agents would desire to do. As he writes:

our ϕ -ing in circumstance C is right if and only if we would desire that we ϕ in C, if we were fully rational, *where ϕ -ing in C is an act of the appropriate substantive kind*: that is, it is an act of the kind picked out in the platitudes about substance. (Smith, 1994, 184, emphasis in original)

The emphasized portion of that quotation is a gesture at a promise that Smith does not fulfill: without an account of substantive rationality, Smith is not in a position to tell anyone what they should do. Nevertheless, the point is clear enough: by thinking subjunctively about the desires of fully rational agents, Smith thinks we can figure out what we should do on any particular occasion. An example will be useful to bring out the features peculiar to Smith's view.

Suppose an individual is trying to figure out what to do. Imagine, for example, that Mary, a genuinely selfish individual with a policy of acting only when it will advance either her immediate or long-term ambitions, is confronted with a practical problem. The airplane she is in has encountered some technical difficulties. It is going to attempt an emergency landing in a nearby river, and everyone has been asked to put on a life-jacket. But the woman sitting next to Mary, who confesses that she does not know how to swim, is without one. Mary, who is an excellent swimmer, understands the situation and must decide what to do. On Smith's account, the way to figure out what it would be right to do is to think about what a fully rational agent would do. If we assume that such an agent would give up her life-jacket in such a situation, Mary should give up her life-jacket.⁹ And, furthermore, if Mary thought

⁹Again, since Smith does not provide a substantive account of rationality, all we can go on to make the case that an individual is acting rationally (or not) are intuitions. There are obvious

subjunctively about a fully rational agent, she would see that she has a normative reason to give up her life-jacket.

Provided certain assumptions about rationality, Smith's portrait provides normative guidance that is different from what would be required of Mary by the other accounts. On Velleman's view, rational agency is acting in order to be intelligible to oneself. Since we have assumed Mary is genuinely selfish, giving up her life-jacket would be unintelligible to her, and as a result, she should not do it. Similarly, for Bratman, an individual has reason to x when x -ing would be endorsed by the pro-attitudes responsible for her cross-temporal organization. Since the central pro-attitude that organizes Mary through time is her selfishness, and since her selfishness could not endorse such an action, she should not give up her life-jacket.

Now, I do not mean these points to be objections to any of these views. Whether Mary *should* give up her life-jacket does not matter to the point I am making. All that matters is that Smith's account suggests a different course of action for Mary than those of Velleman and Bratman. We may cringe at the idea of a person not giving up her life-jacket for such selfish reasons, but we still understand such an action as reasonable—as something that, though cringeworthy, Mary could find reason to do. If all this is right, it looks like Smith's view is not concordant with those of Bratman and Velleman: there are occasions when Smith's agent should do things differently than the agent's portrayed by the latter two theorists.

2.3.2 Deep Desires

But what about the views of Velleman and Bratman? Are their views consonant with each other? To see that they are not, a couple of additional examples will be helpful. Consider Velleman's view first. The opening line of his early book, *Practical Reflection*, is a question. He asks: "what do you see when you look at your face in the mirror?" (1989, 3). The question is significant because it frames the idea that agents must look at themselves in order to make their actions intelligible. It is also informative in that it shows how Velleman thinks about the distance between an

problems with that fact. As a result, I have tried to be as neutral as possible about rationality while still making the examples work.

agent and her sense of agency: as though moving closer to the mirror could move me closer to my sense of self. The view is, I think, peculiar in that sense. A glance in the mirror may not tell me much about myself, but if I really stare, I can discover things about myself that were previously unnoticed. The metaphor is an apt expression of his view.

Suppose, then, that I have a tendency to avoid risk. I prefer to stay away from high places, I prefer not to speed, I do not swim in the ocean, I do not ski, I only ride my bicycle while wearing a helmet, and so on. Acting in accordance with these desires and preferences, I decide to go traveling through Zimbabwe with a friend who is much riskier than I am. During our travels, she decides to go bungee jumping over the Zambezi River. I decide to tag along without intending to participate. At some point, however, I am asked, “Would you like to have a go?” Initially, I refuse (since I know what I prefer), but after a bit of internal reflection I find that deep down I do desire to do something risky, just this once. It is as if I pull the mirror of myself slightly closer and see that, in fact, I do desire to take a significant risk. Although I had been unaware of this feature of myself, now, once I pay attention, I see clearly that I do have such a desire. But I know bungee jumping over the Zambezi River is one of the more unsafe locations to take on such risky behavior. Nevertheless, I believe the desire reflects what I want to do.

Intuitively, such risky behavior might seem irrational in the sense required of Smith’s view.¹⁰ Indeed, if we consider the risky behavior to be something a Smithian agent would not desire to do (because it cannot be derived from the desires of a sufficiently large set of other agents), Smith’s view is going to have a difficult time representing the idea that I should go bungee jumping. Similarly, since the behavior is straightforwardly incompatible with pro-attitudes that organize my agency through time, it is not something that Bratman’s account can portray as a justifiable action. Since my policies, which have been responsible for organizing my agency over time,

¹⁰In other words, taking on such risk might not reflect the widely shared desires that could, on Smith’s account, be used to determine the counterfactual desires of fully rational agents. To avoid confusion, I will use the phrase ‘Smithian agent’ to refer to the counterfactual ideal characteristic of Smith’s view. A Smithian agent, then, is an agent whose counterfactual desires have been established by subjunctively reasoning from the desires of a large enough group of individual agents.

clearly conflict with such a desire, Bratman has no way of making sense of the idea that I should go bungee jumping on this occasion. But on Velleman's account, my desire to do what I understand about myself, and my capacity for understanding myself, provide a normative reason for jumping. Since taking the plunge is the only way to make myself intelligible to myself, and since I have a higher-order desire to do what makes myself intelligible to myself, I *should* jump. If that is right, then, as with Smith's view, Velleman offers a portrait of agency that provides substantively different normative guidance on particular occasions from the other two.

2.3.3 Planning Agents

Another example, which will bring out the idiosyncrasies of Bratman's view, will prove helpful as well. Consider, then, an agent who knows she does not want to *x* because *x*-ing is something a Smithian agent would not want to do. Nevertheless, *x*-ing is an action that is endorsed by pro-attitudes that organize and give authority to her agency over time. For example, Joy has always valued learning, has made it a policy of hers to learn as much as she can in life, has plans of becoming a professional philosopher, and has recently been accepted to graduate school. She knows that state and federal budgets have cut funding to higher education, she knows that these cuts tend to hit disciplines in the humanities the hardest, she knows that the job market for philosophy is over-saturated with qualified people, and she knows that graduate school does not pay much. Indeed, thinking about these facts, she realizes that she is going to be poor for several years to come, that she will likely have a difficult time finding a job, and that even if she does find a job, it will be one that, given funding priorities, may be difficult to keep. But Joy does not want to be poor, she does not want to have a hard time finding employment, and she does not want a potentially unstable job. What should she do?

Again, it does not matter for my purposes what the right answer is, or whether there is a right answer. Rather, all I want to point out is that the different theorists must provide substantively different answers to this question. Assuming that desiring to pursue a discipline that has small (and diminishing) hope for future employment is not a widely shared aim of agents, a Smithian agent would not desire to do such a

thing. As a result, it would be difficult on Smith's account to claim that Joy should go to graduate school. Similarly, if Joy realizes upon deep reflection that she does not desire (or prefer) to be poor, out of work, and in an unstable job, Velleman cannot represent the idea that Joy should go to graduate school. Indeed, on Velleman's account, doing so would be an act of estrangement. Yet, for Bratman, Joy does have justifying reasons for going to graduate school. Since she has organized her agency according to these plans, policies, and values, continuing on with them is part of what it would be for her to continue to persist as the agent she is. Bratman's portrayal of rational agency, then, can prescribe actions for Joy on this occasion that cannot be prescribed by the other two theorists.

The discordance I am emphasizing with these examples should not be particularly surprising. Views that depend on distinct intellectual capacities associated with different pro-attitudes will naturally result in normative principles that diverge in a range of cases. But that divergence suggests something interesting. In particular, it suggests that instrumental agency is not a uniform concept. Since in each case what we are moved to do and what we should do according to the individual views depends upon capacities and pro-attitudes that are different, the guidance afforded by those views varies as well. Yet despite the difference between the normative aspects of these views, most of us, I think, can recognize the reasons provided by each as compelling reasons for action.

2.4 Nonoptional Capacities of Agency

At this point, we have on our plate a collection of views that are normatively incompatible: on particular occasions, what an agent should do according to one view will not necessarily transfer to the others. If that is right, we are faced with a problem: we can either embrace one account while rejecting the others, adopt a strategy that can integrate the incompatible views, or reject them all.

In this section, I am going to argue that we should not accept the first or third options¹¹ by showing that the intellectual capacities and pro-attitudes emphasized by Smith, Velleman, and Bratman are nonoptional in the sense that, if instrumentalist

¹¹I turn to the second option in Section 2.5 of this chapter.

views fail to account for the different capacities and attitudes used by each theorist, we are left with a distorted conception of ourselves as instrumentally rational creatures. This is due to the fact that the capacities and attitudes at the heart of each view are needed to deal with practical issues human agents have an interest in managing. In this sense, each account is nonoptional. If I am right, we are faced with a further puzzle: how should we square the fact that three presumptive theories of instrumental agency appear to be nonoptional, yet result in normative demands that conflict?

2.4.1 Subjunctive Thought

Consider Smith's view first. To dismiss his view in favor of those proffered by either Velleman or Bratman would mean abandoning—or at least consigning to a subordinate role—subjunctive thought as a strategy for determining what one should do. But subjunctive thought serves an essential role in resolving a variety of practical problems, problems that are unlikely to be addressed using the modes of reasoning typified by the views of Velleman and Bratman. In other words, there are cases where asking, “what would a rational agent do?” is the only way to figure out what one should do. Asking whether my desires accord with the plans and policies responsible for my cross-temporal organization, or whether what I want to do is intelligible to me as something that it makes sense for me to do, cannot provide the right answer.

Suppose, for example, that an individual is trying to overcome something of a compulsion, say, an urge for cleanliness. It is not a particularly negative compulsion; the individual cleans mostly when others are not around, he does not do it to excess, and he does not hurt anyone by doing it. In fact, cleaning is something that he really enjoys: it is something that he is happy to do, an aspect of himself with which he is satisfied, something that organizes him and his life over time, and something that, even after deep reflection, he desires to do and is motivated to do in the light of his own self-understanding. In short, his compulsion is, given the views of Velleman and Bratman, something that he should keep doing. But it is still a compulsion.

A representation of rational agency that grounds practical normativity in subjunctive thought is in a unique position to provide practical direction in such a case. It provides a resource for thinking about what to do that does not depend solely on an individual's own motivational attitudes. In this way, an individual compelled to

clean can ask, “What would an individual without such a compulsion do?” and in asking that question, he can find real practical guidance. The value of the mode of thought, then, is to be found in the fact that an individual who recognizes something about himself, something that he enjoys, identifies with, and finds deeply satisfying, can still recognize that it is not a particularly virtuous feature of his own agency. In such a case, he needs a way to think about what he should do—that is, how he should live—that does not rely on his own higher-order attitudes. Subjunctive thought provides the necessary resource.

The point generalizes: since we all have features about ourselves that we like—features that we use to organize our agency over time, things that we enjoy, find satisfying, and believe to reflect desires that we really endorse—yet recognize to be less-than-ideal, we need a way to think about what to do that does not depend on those features. We all have habits, ticks, compulsions, urges, and dispositions that we thoroughly identify with, find satisfying, and enjoy, yet recognize that we need to change. But an account of normativity that is grounded solely in the attitudes of our own agency will lack the resources for thinking about how to change them.

Changing aspects of ourselves that we like because we think they are less-than-ideal is not the only reason we need the capacity for subjunctive thought in our practical repertoire. Consider, in addition, its role in practical instruction—that is, in instruction about how to do things. In order to learn how to perform new tasks, or to learn how to perform tasks better, agents frequently rely on their capacity to imagine and think about what is not the case, or what is only possible in an imagined scenario. Suppose, for example, that an acting student is trying to learn how to convincingly portray the Scarecrow from the *Wizard of Oz* in an upcoming play. To do so, she must think about an imagined scenario: in what way would a creature without a brain behave? Only by reflecting on such an impossible scenario will she learn how to convincingly act like Scarecrow. Of course, this mode of thought is not just used to learn how to play-act. It is also used to do a variety of other things: even top athletes learn to play better by imagining what their heroes would do. Tiger Woods, for example, might try to improve his golf game by imagining what Jack Nicklaus would do. The parallel with Smith’s normative claims is transparent. It

seems, then, that subjunctive thought plays a significant role in determining what one should do in a variety of cases that are important to the practical management of human agency. To give it up would require giving up important aspects of our own lives.

We cannot subordinate the normative role of subjunctive thought to the accounts put forward by Velleman or Bratman either. Since their accounts rely solely on an individual's own sense of agency to provide normative guidance, their accounts are unable to make sense of an individual setting out to make changes to his agency if he is satisfied with them. If we know who we are and are satisfied with who we are, there is no reason, on the views of Bratman and Velleman, to change.¹² To make such changes, we need the kind of practical normativity that we get from counterfactual thinking about other, more ideal agents.

2.4.2 Self-Intelligibility

So there seem to be good reasons for endorsing Smith's commitment to subjunctive thought as the means for determining one's normative reasons for action. But the capacity and attitude at the heart of Velleman's view seem necessary too. Indeed, as with Smith's account, if we give up on acting to be intelligible to ourselves, we give up on an essential feature of our own agency, one that is needed to manage practical problems we all (at one time or another) have an interest in managing. These problems, I want to suggest, cannot be managed using the views of Smith or Bratman. In other words, sometimes it is only by asking "how can I act to make myself intelligible to myself?" that we can figure out what to do.

To make the case that the heart of Velleman's view is needed to forestall a distorted view of ourselves as instrumental agents, consider the type of practical problem it is particularly adept at addressing. For nearly all of us, there comes a time in life when we must figure out what we want to do with ourselves by reflecting on our desires and preferences. Frequently, such self-reflection marks a sharp break in the direction

¹²One might suggest that a Bratmanian agent could adopt, as a cross-temporally organizing policy, the practical mode of thought characteristic of Smith's view. In this way, the normative role that such thinking plays in Smith's account could be subordinated to Bratman's view. The worry is a version of the integration problem that I have been putting off for some time now. I will return to that problem soon.

and fortunes of our lives, because it reveals aspects of our agency that are at odds with the aims, ambitions, and hopes that others have instilled in us. Despite the best intentions of our mothers, we do not all hope to become (as Willie Nelson notes) doctors and lawyers and such. Indeed, after reflecting on what we want, many of us are compelled to abandon the plans, hopes, and ambitions that have to that point defined us. This is due, I take it, to the fact that we recognize that such plans, hopes, and ambitions are incongruous with our own preferences and desires. Similarly, when we are trying to figure out what to do with our lives, reflecting on the counterfactual desires of a Smithian agent is of no help. We do not want to know what some hypothetical collection of counterfactual desires would do, we want to know what we, as individuals, should do. To figure that out, we have to determine our actual desires and preferences about how to live by looking inward.

The idea is a familiar one. Consider, for example, a college dropout who abandons his plans to become a lawyer in order to make it big as an artist in New York City.¹³ Such a decision, we can assume, is motivated by deep reflection and the type of soul searching typical of a college student trying to figure out which of his desires and preferences really reflect his actual desires and preferences about how to live. After reflection, he acts in order to be intelligible to himself. In other words, he adopts a course of living that, to anyone other than himself, seems to be both out of touch with who he has been and out of line with the counterfactual desires of Smithian agents, yet he seems justified in doing so. If he did not go to New York after figuring out what he really desires and prefers to do, he would betray himself—that is, his actions would be unintelligible to him as *his* actions.

Of course, it is not difficult to imagine situations of this type—situations where an individual acts intelligibly to himself by acting spontaneously on the desires and preferences he discovers after deep self-reflection—and the reason such situations are not difficult to imagine is that they are familiar. Velleman’s view, of course, has

¹³The story I am telling parallels one told by Bernard Williams in “Moral Luck” (1981b). Williams uses the story for entirely different purposes, however. My point, of course, is not about the role of luck in the lives of agents, but, rather, about the need of instrumentalist views to account for agents that act in order to be intelligible to themselves. I assume that what makes that need compelling is an intuition similar to the one used by Williams to drive his discussion of Gauguin.

the resources to make sense of the actions typified by such cases as actions that the agent should perform. The other views are not in such a comfortable position. Human agents who are committed to their past policies, plans, and values as definitive aspects of who they *really* are will have difficulty making sense of such actions as actions they should perform. Similarly, someone who acts only on the presumed counterfactual desires of a Smithian agent will have difficulty making sense of the idea that one should spontaneously go to New York, since it is unlikely that such counterfactual desires could be derived from the collection of desires typical of others. Such views, then, do not provide a way to make sense of acting in an effort to be true to ourselves in the way Velleman's view does.

The example may seem slightly unusual, but to the extent that it is, it is only to make a point. Instrumentalist views of rational agency need to make sense of the capacity for individuals to justifiably do things that are out of line with the plans, policies, and goals that have sustained and organized their lives. Conversion to a new cause, or to a new mode of life, after engaging in deep personal reflection is a substantive event in the lives of human agents, and theories that fail to make sense of such events distort the way we understand the reasons we have for doing the things we do. Sometimes we feel compelled to give up on our planned lives; sometimes the only way to be intelligible to ourselves (after taking a closer look at our actual desires and preferences) is to move to New York in order to be an artist. The mode of thought leading to such actions cannot be represented as being authored by a Bratmanian agent, and, similarly, if our self-intelligible actions are motivated by desires and preferences that are out of line with the counterfactual desires of a Smithian agent (and it is not a stretch to assume that they frequently are), they cannot be represented in that way either.

It is precisely because deep reflection on what we really want often promotes actions that we feel compelled to perform that ignoring that aspect of our agency would distort our understanding of instrumental human agency. Indeed, I take that aspect of Velleman's view to be one of its most appealing features. It is committed to the value of self-knowledge and the importance of acting in accordance with what one knows about oneself. To give up on the mode of reasoning, and the actions that

stem from it, is to abandon something of deep practical importance to the lives of human agents. Consequently, we need the capacities at the heart of Velleman's view. Sometimes the only way for an individual to control his own life is to figure out what he really believes about his desires and preferences, and to act in an effort to make himself more intelligible to himself.

2.4.3 Temporal Organization

As with the accounts of Smith and Velleman, so too with Bratman's: understanding ourselves as instrumental agents requires incorporating the heart of his view. Without it, we give up on representing aspects of human agency important to our practical lives. This way of thinking seems to be endorsed by Bratman himself. His early work rested on the idea that we would be different kinds of creatures if the capacities and attitudes at the heart of his view were not part of our cognitive makeup (Bratman, 1987), and these ideas have been expanded in *Structures of Agency* to include values, policies, and other higher-order pro-attitudes. If we try to abandon these features of our agency, or subordinate them to the alternatives of Smith or Velleman, we abandon part of our practical repertoire, which is needed to manage our lives as human agents.

Again, to see this, consider the kind of practical problems that a Bratmanian view of agency is particularly adept at addressing. It is difficult to deny the importance of values, plans, and policies when explaining and justifying the things we do for reasons. Indeed, some of the things we do seem to be done solely because of the plans, policies, and goals we possess. Suppose, for example, that an individual possesses a self-governing policy to be temperate and sober. His father was an alcoholic, so he has been committed to not drinking his entire life, and, in fact, the policy is an aspect of his agency with which he is satisfied. Yet suppose that one morning, such an agent is asked by a group of coworkers to go out for a drink that evening. After an afternoon of deep reflection on what he *really* desires and prefers, he realize that he really does, deep down, want to go out drinking with his coworkers. And in fact, he also believes that going out for a drink would be what a Smithian agent would do. After all, the boss will be there, and he will be able to chat with him about his latest project. As a

result, the night promises to help his career. (Certainly, the desire to help one's career must be, if anything is, contained in the counterfactual desires of a Smithian agent.) Nevertheless, when the time comes, he turns down the offer. When asked why he does not want to go out, he responds (maybe a little stiffly), "Because I value temperance and have a policy against alcohol consumption." The response, I take it, explains and justifies his action. Appealing to the temporally persisting policies and values with which we identify seems straightforwardly reasonable, and asking someone to act against them, even if such an action is something a Smithian agent would do, or something you know the individual really, deep down, desires to do, is an affront. It is asking them to give up on who they have been. If that is right, instrumentalist views need to account for this aspect of human agency when accounting for our instrumental rationality, and it does not look like Smith and Velleman are in a position to do so.

But such actions, and the reasons that motivate them, might, with a little twisting and turning, still be explained by Velleman and Smith. In order to see the peculiarity of Bratman's account, then, we need an example that runs contrary to what an agent, after due reflection, prefers to do, and which also is not something a Smithian agent would desire to do. Consider, then, the family traditions to which many of us are beholden, traditions which express the values, policies, and plans peculiar to ourselves and the families to which we belong. These familial attitudes often play important roles in holding us together as individuals and as members of particular family units. I am thinking, for example, of the traditions that guide actions solely on the basis of their being traditions of the Mosdell family, or the White family, or the Sawyer family, or whatever. Being committed to such traditions, and performing their associated actions whenever they come up, often serves to define us as members of the particular families to which we belong—that is, the performance of such traditional actions are essential for understanding ourselves as Mosdells, Whites, Sawyers, or whatever.

Consider the actions that stem from such traditions, actions that are not motivated by reflecting on the counterfactual desires of a Smithian agent, and do not serve to satisfy any desire or preference of the individual. (For example, the tradition may involve baking a Thanksgiving turkey using a peculiar sort of homemade stuffing that no one really seems to enjoy.) That last condition may seem like a sticking point.

After all, if an agent does not really desire or prefer to do x , why should he do it? The answer, I am assuming, is because he is a Mosdell, or a White, or a Sawyer, or whatever. In fact, despite the fact that we often prefer and desire not to perform the actions associated with such traditions, we do them anyway because, in some deep way, they are tied to the temporally extended nature of our being members of the family with which we identify. My point, then, is that human agents feel compelled to perform actions associated with any number of traditions, not because they really prefer and desire to perform the relevant actions, and not because the actions should be performed given the counterfactual desires of a Smithian agent, but, rather, because agents identify with the families and traditions that hold them together as temporally persisting members of particular families.

In the face of such an observation, one might simply claim that in doing such things, we are simply acting irrationally. Well, maybe. After all, if we do not like performing such actions, if the performance of certain aspects of some traditions are unpleasant, undesirable, or otherwise disagreeable, why perform them? The point, however, is meant to be agnostic about the substance of rationality, and, instead, is designed to simply emphasize that being moved to perform traditional actions plays a substantive role in the lives of human agents. To the extent that that role is tied to holding us together as members of particular family units, instrumentalism needs to make sense of the actions that stem from it as something human agents can and should do. To act from tradition simply because it is traditional to do so is sometimes the only way to figure out what we should do. If that is right, Bratman's view provides a solution to a set of practical issues that most of us have an interest in managing, and if we give up on that view, we give up on a valuable resource for explaining and justifying such behavior.

2.4.4 Nonoptional Views

To recapitulate: I have tried to show in this section that the heart of the accounts put forward by Smith, Velleman, and Bratman are nonoptional when accounting for human agents as instrumentally rational. We cannot give them up without giving up important resources for explaining how we deal with practical issues we all seem to

have an interest in managing. If that is right, the capacities and attitudes central to each account are necessary for making sense of ourselves as instrumentally rational agents. Yet as we saw in the previous section, the normative guidance derived from each of these accounts conflicts on particular occasions. While one theorist's account justifies doing x on a particular occasion, the accounts of the other two prescribe doing not- x on the same occasion. But surely, accounts of agency that prescribe conflicting actions on the same occasion cannot each be correct. As a result, we are faced with a puzzle: how are we to square the fact that 1) the core aspects of three presumptive theories appear to be nonoptional, and 2) the prescriptive guidance they offer is incompatible?

2.5 Models of Agency

One strategy for dealing with this puzzle is to move to a more general conception of agency. By formulating the rational grounds of instrumental action without relying on any *particular* feature of human agency, one might hope to formulate a normative principle for acting that is more general. Such an approach, however, divorces these views from their normative implications. This observation, I take it, is what motivated these theorists to move away from the general instrumentalist conception of rational agency proposed by Williams. An alternative strategy is to integrate. But why should we think that the grounds of instrumental agency can be integrated into a single, unified principle of agency? I do not think we should.

2.5.1 Against Integration

There are, I think, several reasons to believe that we cannot integrate these instrumentalist views into a single, unified account. Before beginning in on those arguments, however, let me emphasize that my aim in this chapter has been to understand the motivational attitudes and rational capacities at the heart of *human* agency. The emphasis is meant to contrast the aim of understanding ourselves with the aim of understanding some ideal notion of rational agency. It is because of the deep complexity of our own agency that I think an integrated view of instrumental agency is unlikely to be in the offing. When trying to make sense of the relation between what actually motivates creatures like you and I to act and what should

move us to act, there is a gap that needs to be filled. I have tried to show that that gap cannot be filled with a single view. We need to account for the different attitudes and capacities of actual human agents in order to understand our nature as instrumentally rational creatures. To that end, I have argued that each of the views we have surveyed is nonoptional. I have also argued that these views are normatively incompatible in a range of cases. And the question before us now is, can this incompatibility be overcome by integrating the different views into a single, unified principle of instrumental agency?

Of the accounts we have looked at, the most likely prospect for integrating the others to it is Bratman's, because the higher-order self-governing policies characteristic of his account are *very* flexible. (Recall that such policies are second-order attitudes that, to the extent that agents are satisfied with them, provide reasons for treating a lower-order desire as an end to be pursued.) Given such liberal constraints on the notion of a self-governing policy, why not simply integrate the views of Smith and Velleman into Bratman's? In other words, why cannot an individual possess a self-governing policy to act when he believes that his desires align with the subjunctively derived desires of a Smithian agent. Or, for that matter, why cannot an agent possess a self-governing policy to act only when he has determined what he believes his desires and preferences to be? I want to suggest that it is because instrumental agency looks to be at best a patchwork of disjoint models, and each of which is needed to represent human agents as instrumentally rational.

If the alternative views we have surveyed in this chapter can be integrated into a single, overarching view of agency, the psychological attitudes and intellectual capacities at the heart of each will have to be unified around one or another capacity or attitude.¹⁴ If the different capacities and pro-attitudes that motivate action work against each other, it will not be possible to coherently integrate them. As I suggested, Bratman's view holds the most promise for integration. Is it possible, then, to bring together the attitudes and capacities at the heart of the views of Smith or Velleman

¹⁴Of course, the type of unification I am discussing is unlikely to be possible with more than one agent. Consequently, integrating the psychic economy of more than one agent in an effort to provide a single principle of instrumental agency, which applies to all human agents indifferently, seems to be an unlikely prospect.

with Bratmanian self-governing policies?

To answer that question, we need to understand what is required for integration, or unity, of this type. Bratman follows Harry Frankfurt¹⁵ in suggesting that the unity of an individual's psychological attitudes depends on his being satisfied with the self-governing policies that direct his agency. To be satisfied in Bratman's sense, however, means that any particular policy with which one identifies is not incompatible with, or challenged by, other self-governing policies.¹⁶ Of course, being satisfied in this sense cannot occur if an agent adopts both Velleman's higher-order desire to act on self-evaluative beliefs and a policy to act in line with the counterfactual desires of a Smithian agent. As we have seen, if an individual tried to adopt both views as self-governing policies, it would result in normative tension. The two views conflict in a range of cases, and, consequently, an agent who tried to act in accord with both would eventually find himself internally at odds with himself. Trying to realize a Smithian agent's counterfactual desires and trying to be true to one's own desires and preferences would, over time, pull an individual in two incompatible directions.

But what about incorporating either Smith or Velleman into Bratman's view? Is it possible to adopt a self-governing policy to be either a Smithian agent or an agent who has a higher-order desire to act on self-evaluative beliefs about what he desires and prefers? I doubt it. Consider, first, trying to merge Velleman's view with Bratman's. As we have seen, it is crucial to Bratman's view that we are temporally extended creatures. Indeed, it is at the very heart of his account that self-governing policies with which an agent is satisfied serve to organize and sustain the Lockean ties necessary for an agent's temporal persistence as a single individual. But as we have seen, the view has difficulty making sense of spontaneity. It is for this reason that,

¹⁵In "The Faintest Passion," Frankfurt writes that "the essence of rationality is to be consistent; and being consistent ... means proceeding so as not to defeat oneself" (Frankfurt, 1992, 7). According to Frankfurt's view, an agent whose higher-order desires are not at odds with each other, one who is, to use Frankfurt's terminology, not ambivalent is "wholehearted in his higher-order attitudes and inclinations, in his preferences and decisions, and in other movements of his will" (1992, 7). Such an agent is, again to use Frankfurt's language, satisfied with who he is, and has no interest in changing his agency, because the attitudes that make it up are not internally at odds with each other.

¹⁶See Section 2.2.3 of this chapter for details.

for example, Bratman's view struggles to explain the decision to go bungee jumping if such a decision is out of line with an agent's risk-avoidance policies. Of course, such spontaneity is not a problem for Velleman's view, or so I have argued. As long as an agent acts on his beliefs about what his actual desires and preferences are, he acts rightly. Similarly, Velleman's view cannot make sense of an individual being motivated to act purely from tradition. If an individual possesses no desires or preferences to perform some action—that is, if his action is unintelligible to him as something he would like to do—yet feels compelled to do because it is traditional, Velleman's view cannot make sense of it as something the agent should do. Attempting to merge the views of Velleman and Bratman, then, results in principles of action that are at odds. Consequently, an agent could not reasonably adopt both while aspiring to be an integrated agent.

Ignoring that difficulty for a moment, what about adopting a self-governing policy to act like a Smithian agent? Is integrating the Smithian view with Bratman's possible? Again, I do not think so. The problem is not that such an agent is conceptually impossible;¹⁷ rather, the problem is that adopting a self-governing policy to act like a Smithian agent requires a type of cognitive dissonance, since it requires trying to govern one's actions with two disparate psychological perspectives. To see this, suppose we try to integrate Smith's view of rational agency with Bratman's. The resulting self-governing policy would be to act on one's subjunctively determined beliefs about the counterfactual desires of purely rational agents. Aside from the fact that we do not have any idea what those counterfactual desires might be, adopting the Smithian view as a higher-order self-governing policy would mean adopting a policy to act as one believes an ideal agent would act. But committing to such a policy is tantamount to adopting a policy to act in a manner that is essentially beyond human capabilities. Human agents cannot reasonably aspire to act like ideal agents because human agents cannot, in the face of our cognitive limitations, acquire the psychic unity necessary for being an ideal agent. Consequently, adopting a higher-order policy that says to act as though one has become an ideal agent is tantamount to adopting a

¹⁷Whether it is conceptually possible to merge Bratman's view with Smith's, however, depends on Smith's substantive account of rationality, which he does not provide.

policy to pretend that the perspective of a form of agency that is not possible for human agents can move one to act.

A different way of framing the complaint I am making here is to say that Bratmanian agency essentially requires looking inward—looking, that is, toward the motivational attitudes that hold us together as agents—for normative guidance. Smithian agency, on the other hand, requires looking outside of ourselves for such guidance. To act as a Smithian agent, an individual does not ultimately look to his desires to figure out what to do, he looks to the hypothetical desires of a purely rational agent. But it is just the opposite for Bratman: to figure out what I should do, I look to the self-governing policies that hold me together as the individual I have been and aim to continue to be. Now, although it is conceptually possible that those two perspectives could be integrated—if somehow it turned out that the attitudes that hold me together just were the attitudes of an ideally rational agent—the fact that we are mere human beings seems to preclude that possibility. It is not very likely that my attitudes and desires will just turn out to be the attitudes and desires of the conceptual ideal. To try to adopt both perspectives as grounds for rational action, then, demands a type of cognitive dissonance. It requires looking for normative guidance by looking both inward and outward at the same time. To the extent that these two perspectives point in different directions, attempting to act in harmony with both will pull one in conflicting directions. Of course, since the normative grounds of Bratman's view depend on a sufficiently high degree of psychic unity, trying to adopt a policy that leads to this sort of cognitive dissonance would undermine the normative basis of his view.

None of this is meant to suggest that we cannot have a variety of different policies that are each tied to different aspects of our own agency. It may be that we have policies that are deeply incongruous. Nor am I trying to suggest that Bratmanian policies cannot accommodate a variety of potentially discordant self-governing policies. It may be that we do act according to policies with which we are not fully satisfied but which, nevertheless, serve to more or less hold us together through time. Be that as it may, I want to insist that to the extent that the self-governing policies with which we identify are not fully compatible, we cannot integrate them into a

single, overarching view of agency. If we require policies and attitudes that are not fully harmonious in order to deal with practical problems we all, as human agents, have a stake in managing, when we act to satisfy one of them at the expense of the others, we inevitably act in a way that frustrates aspects of ourselves with which we identify.

In an effort to provide a unified principle of instrumental agency, self-governing policies sound like a good idea. But, in fact, they do not serve to provide the necessary unity for making sense of instrumental agency as it is found in human agents. Instead, they serve at best to conceal a number of different and conflicting models of instrumentally rational agency. In fact, we see this by comparing the different views side by side: to the extent that making sense of both our spontaneity and our commitment to tradition are required for understanding ourselves as human agents, we have to give up on the idea that we can integrate the presumptive theories that try to make sense of such actions. Similarly, to the extent that making sense of our nature as instrumental agents depends on actions that rely on counterfactual conceptions of ideal agency and our own psychic unity, we will have to abandon the notion that we can integrate the different views of agency. When we look closer, instrumental agency is not a single thing. It is a patchwork of disjoint models that represent different capacities and attitudes required for understanding ourselves as the rational creatures we seem to be.

2.5.2 Facades of Representation

That we cannot integrate different views of instrumental agency should not be surprising. Indeed, consider that unifying principles that explain complex objects are something of a rarity, and are often only described by a lattice of models.¹⁸ A couple of examples, one from Nancy Cartwright and one from Mark Wilson, will help to make this point. In *How the Laws of Physics Lie*, Nancy Cartwright argues that the general laws of physics do not state the truth. To support this claim, she argues that there is a trade-off between truth and explanation. For general laws to describe the behavior of particular objects, they must ignore the variety of composite causes

¹⁸See, for example, Ravetz (1971).

that go into explaining what the objects actually do. But when composite causes are taken into account in order to explain an object's behavior, general laws fail to be true of it. In other words, since the general laws of physics ignore complexity, they do not explain how things actually behave. Their truth, then, is had at the cost of explanation. As Cartwright puts it:

The laws of physics, . . . to the extent that they are true, do not explain much. We could know all the true laws of nature, and still not know how to explain composite cases. Explanation must rely on something other than law. (1983, 72-73)

Indeed, to explain complex phenomena, physicists use models, which represent certain aspects of a things behavior while ignoring others. Again, from Cartwright:

We construct different models for different purposes, with different equations to describe them. Which is the right model, which the true set of equations? The question is a mistake. One model brings out some aspects of the phenomenon; a different model brings out others. Some equations give a rougher estimate for a quantity of interest, but are easier to solve. No single model serves all purposes best. (1983, 11)

The point, then, is that due to their simplicity, general laws do not actually represent how things behave. To do that—that is, to actually represent how physical objects behave—we need a variety of models designed for different purposes, each of which is designed to explain different aspects of the behavior in question.

Mark Wilson makes related points in *Wandering Significance* (2006). There he argues that the meanings of linguistic predicates wander for a variety of reasons. Only with considerable effort can we hold them constant. Indeed, one reason meanings wander is that we overlook the more complex behaviors hiding behind their seemingly straightforward surface meanings. Consider an example that Wilson develops at length. The predicate 'light intensity' is used to describe how incoming light scatters at the tip of a completely reflective razor blade. It represents a physical quantity that seems to correlate with a single analytic function. But, as it turns out, the analytic function used to characterize what 'light intensity' means is not unitary. Rather, it covers a patchwork of facades—or fragments of representation that correspond to different analytic functions—that are used to determine 'light intensity' for different segments of the razor's tip. These facades have, to use an apt metaphor, cracks between them—that is, regions of 'light intensity' that are too complex to be worth

calculating. To get around the cracks, theorists ignore them, moving abruptly between facades in order to calculate ‘light intensity’ for well-defined segments around the razor. The different analytic functions that go into distinguishing each facade are then, through brute force, welded together to “provide a reasonable facsimile of the way light actually scatters around a razor blade (2006, 322). In other words, the complexity involved with understanding the predicate ‘light intensity’ is generally ignored, giving its meaning the appearance of relying on a seemingly unitary analytic function. What the predicate actually means, however, is more complex.

The examples developed by Cartwright and Wilson are not unusual: often seemingly simple principles conceal hidden complexity that does not nicely conform to the overarching idea espoused in the principle. As Cartwright says, “things are made to look the same only when we fail to examine them too closely” (Cartwright, 1983, 19). These points apply to Bratmanian policies and to the question of determining the grounds of our own instrumental rationality. As we have seen, the capacities and pro-attitudes needed to explain an individual’s instrumental reasons are unlikely to be represented with a single theory. If that is right, it looks like trying to understand the grounds of our own instrumental agency is like trying to understand ‘light intensity’ or the behavior of other physical phenomena. Broad principles fail to reflect the complexity of the phenomena, and when we dig deeper, what we find is not a unitary principle, but rather, a patchwork of models stitched together by their common aim to explain some complex subject matter.

2.5.3 Disunified Models

The views of Smith, Velleman, and Bratman depend on capacities and pro-attitudes that we cannot do without if we hope to explain the range of instrumental reasons that move human agents to act. Yet there are obvious conflicts between the portrayals of agency proffered by each theorist. However, rather than abandoning the tools needed to explain ourselves and the capacities and attitudes that move us to act, or trying to formulate a unitary principle that can represent everything we do at the cost of explaining it, we should adopt the stance of a modeler: represent specific aspects of rational human agency by using a variety of models.

2.6 The Disunity of Instrumentalism

Let me wrap this chapter up by pulling the threads together. Accounts of agency aim to explain what human agents *are* motivated to do as well as what they *should* be motivated to do. Smith, Velleman, and Bratman provide portrayals of agency with both aspects of explanation. Yet the normative guidance they provide conflicts in a range of cases. Despite these conflicts, however, I have argued that the features used to develop their accounts should not be ignored. The result is a puzzle: how do we square the fact that three different accounts provide different prescriptions for action while acknowledging that the capacities and attitudes used to derive their prescriptive content cannot be ignored? In response to that puzzle, I have argued that these discordant representations of agency should be understood as models, or localized patches of representation that portray certain aspects of end-oriented action, but not others.

But there are two problems that seem to result from this conclusion. First, since the normative components of the different models conflict, they cannot be conjoined to form a nice overarching theory. This means that the idea of instrumentally rational agency is, in a certain sense, disunified: what we mean by it cannot be represented with a single, normatively compelling principle of action. Rather, as we have seen, to be instrumentally rational means to be moved by normative reasons grounded in one's cross-temporal organization, self-intelligibility, and subjunctive thought. Of course, we may need other models as well, but if I am right, at least these three will be needed to account for the variety of reasons we can and should be moved by. But this need not be alarming. Complex phenomena—such as our own agency as rational creatures—are rarely explained by simple, unified principles. And if that fact has not impeded our capacity to represent and explain the natural world, there is little reason to think that it will hinder attempts to represent and explain ourselves.

Nevertheless, there is an additional problem to worry about. If we give up on the idea of finding a single, unified account of instrumentally rational agency, one that can provide universal normative reasons for acting, do we not need some other principle to determine when one model is to be preferred over another? Well, that depends. I have suggested that general principles are hard to come by and agreed

with Cartwright and Wilson that when they are found, they often do not explain much. If that is right, we should expect a general principle aimed at determining when one model should be used over another to meet with the same problem. This fact undermines attempts to provide a single, unified account of the things we do for reasons, and, instead, leaves us with a concept of agency that is importantly disunified.

CHAPTER 3

MODELING EXPRESSIVE ACTIONS

3.1 Introduction

In “A Plea for Excuses,” J.L. Austin warns that representing the variety of things we do with a single model may distort the finer filagree of human action. He writes:

We take some *very simple action*, like shoving a stone, usually as done by and viewed by oneself, and use *this*, with the features distinguishable in it, as our model in terms of which to talk about other actions and events: and we continue to do so, scarcely realizing it, even when these other actions are pretty remote and perhaps much more interesting to us in their own right than the acts originally used in constructing the model ever were, and even when the model is really distorting the facts rather than helping us to observe them. (1961, 150, emphasis in original)

The sense of ‘model’ being used by Austin is not entirely clear,¹ but the warning is: relying on simple representations of action may lead us to overlook, ignore, or distort alternatives that may be just as philosophically interesting. Indeed, despite Austin’s warning, contemporary philosophers tend to represent action using only a handful of very simple models. According to one such model, whenever an agent performs intelligent actions—that is, actions done on purpose, for reasons—he knows what he is doing and why he is doing it. Central to this conception of action is the idea that rational agency—or the capacity to act for reasons—presupposes practical knowledge, or an understanding of how to do the things one is in the process of doing.² In this

¹For my purposes, it does not matter what Austin means by ‘model’, but since the term will be used throughout this chapter, let me remind the reader of its sense. Models represent by accurately portraying some of an object’s features while distorting others. Since models contain distortions, to represent every aspect of a complex object with models, more than one model is required. For more details on the sense of ‘model’ being used, see the introductory chapter.

²Knowing what one is doing while doing it is what I called practical knowledge in the last chapter. My chief aim in this chapter is to challenge the idea that such knowledge is required for intelligent action. ‘Practical knowledge’ is distinct from ‘knowledge-how’. Indeed, the latter, which is typically used to characterize a type of knowing that is different from knowledge-that, is used to

chapter, I want to challenge that model of action by identifying a class of actions that are, I will argue, done on purpose, but without practical knowledge.

To do so, I am going to focus on a class of actions that I will call *expressive actions*.³ Characteristic of them is that they express or reveal an agent's feelings or emotions. I begin, however, by articulating the Anscombian view I aim to challenge. Once that view is on the table, I use recent work by Candace Vogler as an Anscombian foil for my alternative. What I hope to show is that the simple model of action explanation adopted by Vogler fails to represent expressive actions—at least it fails to do so without considerable distortion. If I am right, this class of actions stands in need of an alternative model, one that can represent thoughtful forms of emotional expression.⁴

3.2 Describing Intentional Action

The Anscombian view of rational agency is unfriendly to the more familiar philosophical tradition. According to that tradition, beliefs and desires work together, causing agents to do things for reasons. Desires provide ends, goals, or aims to be

mark dispositional capacities, or capacities to do things. For example, the capacity of insects to walk is typical of knowledge-how: it is knowing how to do something without knowing any potentially corresponding truths about how to do it. In contrast, the term 'practical knowledge' is used to distinguish the awareness an agent has of what he is doing when he is doing something on purpose. In this sense, 'practical knowledge' is had only by rational agents. When an individual has practical knowledge, he is aware of his action and can demonstrate that fact by articulating what he is doing and what he hopes to accomplish by doing it.

³I use the phrase 'expressive action' (and, from time to time, 'emotional action') to refer to actions that are explained in the first instance by appealing to emotions. So, for example, "Why are you punching your locker?" might be answered by "Because I'm angry." In this instance, the action gives expression to an emotion and is an expressive action. In contrast, I will use 'end-directed action' to mark actions explained in the first instance by appealing to an end the individual is trying to accomplish. So, for example, "Why are you punching the locker?" might be answered by "Because my locker is stuck and punching it tends to get it open." In this instance, the action is an end-directed action. More frequently, however, I will simply use the term 'action' and let the context determine the type of action. All of this is meant to avoid the ambiguities, implications, and technicalities of the more familiar, and widely used phrase 'intentional action'. Finally, I should mention that this distinction is closely related to one made by Rosalind Hursthouse in "Arational Actions" (1991). Indeed, the distinction depends on taking seriously the arguments of that paper.

⁴It is important that such actions are more than mere emotional outbursts. Unlike the example from the previous footnote, the type of expressive actions with which I will be concerned are constrained and controlled by an agent in the act of emotional expression. In contrast to punching one's locker, then, an emotion like anger might be thoughtfully expressed by going for a run or, more pensively, by writing a poem.

achieved, beliefs show what is required for realizing them, and when combined in just the right sort of way, they cause individuals to act. As this tradition would have it, the right way to investigate the knot of problems associated with intelligent action is to determine the causal antecedents of rational behavior.⁵

Anscombe, however, thought that such an approach could not get off the ground until the notion of causality became much clearer. As she writes, “the topic of causality is in a state of too great confusion; all we know is that [acting] is one of the places where we do use the word ‘cause’ ” (2000, 10). That Anscombe thought the notion of causality too confused to explain how mental states cause action, however, is not to suggest that she was not concerned with causes. Indeed, the idea that “practical knowledge is the cause of what it understands” (2000, 87) is a central thesis of the Anscombian view. Despite that fact, however, Anscombe and her recent followers⁶ are not concerned with the causal question as it is usually understood. They are not trying to determine the causal *antecedents* of action, but instead, the causal *unity* of thought and action, where that means that what an agent is thinking is what he is doing.

But what does suggesting that Anscombians are trying to understand the causal unity of thought and action really mean? The short answer is that they are attempting to articulate what thought and action must share in order to be unified in rational agency, and they argue that the shared component is a particular logical relation typified by intention. When agents do things on purpose for reasons, the form of description used to articulate what they are doing reveals the logical order shared by thought and action. Agents that rely on thought to govern what they do describe their actions according to a particular pattern shared by both thought and action. In

⁵The Anscombian view of action is unfriendly to the hierarchical views of agency we looked at in the previous chapter as well. Nevertheless, both Anscombian and hierarchical views require that agents who act intelligently have practical knowledge. It is this similarity that I aim to be arguing against in this chapter.

⁶Recently, there has been surge of interest in the philosophical work of Anscombe. Prominent philosophers have taken up her way of speaking about intention in an effort to expand, clarify, and deepen the insights first propounded by her in *Intention*. I have in mind philosophers like Candace Vogler, Michael Thompson, and Sebastian Rödl. But there are others. See especially the recent collection of papers in *Essays on Anscombe's Intention* (2011) for philosophers who have taken up the torch.

terms more familiar to Anscombians: the unity of thought and action is had through intention.⁷ Because acting on purpose presupposes practical knowledge—that is, the capacity of an agent to understand, without looking, what he is doing—when an agent does something intentionally, he can describe what he knows about what he is doing according to a particular form of representation. In such cases, what the agent does and what the thoughts that describe what he is doing share is a specific formal structure.

That, however, is a very short answer. To make sense of it, I need to say more about the notion of ‘intention’, its relation to practical knowledge, and how those two ideas fit with the view that thought and action are unified. Anscombe writes:

the term ‘intention’ has reference to a *form* of description of events. What is essential to this form is displayed by the results of our enquiries into the question ‘Why?’ Events are typically described in this form when ‘in order to’ or ‘because’ (in one sense) is attached to their descriptions. (2000, 85)

When things are done intentionally in this sense, the way actions get articulated conforms to a particular pattern, which is revealed by the phrases ‘in order to’ or ‘because’. This pattern is shared by thought and action and is what marks it off as end-directed action. For example, when my niece asks me why I am grading a stack of exams, and I respond that I am doing so *in order to* figure out final grades for my course, my response describes the order of action and shows that my thoughts conform to it. Since calculating final grades depends on exam grades, I must grade the latter to determine the former. When my thought and action share this order and are unified by it, the thought represented is the action performed. ‘Intention’, then, is the form of description that unifies what an agent thinks and what he does.

⁷I need to say a bit more about terminology here. ‘Intention’ is, for Anscombians, a term of art. It refers to the form of description used to describe all actions that are done on purpose for reasons. I disagree, however, with the claim that all ‘intentional’ actions can be described using that form of description. There are some actions that are done on purpose, which cannot be articulated with the form of description to which Anscombians use the term ‘intention’ to refer. To avoid confusion, then, I will replace that technical term with my own, more neutral phrase, ‘end-directed’. I will only use ‘end-directed’ to refer to thoughts or actions that share the structure typically reserved, at least by Anscombians, for intentional actions. When speaking more generally of actions that are done on purpose (a set that includes, but is not limited to, end-directed actions), I will speak of ‘intelligent actions’. That phrase is meant to be neutral about the form of description appropriate for describing actions done on purpose while, at the same time, emphasizing the connection between what agents think and what they do.

Since the order that unifies thought and action is, for Anscombian, an order of explanation, for the two to be unified, an agent must have the power to understand what he is doing—that is, he must have practical knowledge. It is only by understanding how one’s actions fit together in the service of one’s aims—which may include specific ends, more complex actions, life plans, and so on—that individuals are capable of acting intentionally. Again, from Anscombe:

‘Intentional action’ always presupposes what might be called ‘knowing one’s way about’ the matters described in the description under which an action can be called intentional, and this knowledge is exercised in the action and is practical knowledge. (2000, 89)

An agent who acts intentionally is able to provide a description of what he is doing that places his action within an end-directed order. If an agent is unable to articulate what he is doing according to the relevant form of description, whatever it is the individual is doing is not being done on purpose for a reason.

Now, the type of descriptive awareness required for intelligent action does not depend on an agent’s observations about what he is doing. So, for example, if an individual is opening a bottle of soda, he can describe what he is doing, and thereby know what he is doing, without looking to see that his hands are grasping the cap, that they are rotating, that his arm is moving, that he is flexing certain muscles, or any other observations. What he knows *is* what he is doing because it is being done as end-directed action: his thought and action are unified by a logical order independent of observation. This capacity to know without looking at what one is doing by representing in thought what is being done in action is practical knowledge, which “is the ‘cause of what it understands’ ” (2000, 87). Acting with practical knowledge, then, is what distinguishes end-directed action, or actions done with thought, from other, non-end-directed actions.

We might wonder, however, what the content of an agent’s knowledge must be if he is to act in this manner. What must an agent know when his actions are done on purpose, for reasons? There are an indefinite number of answers to that question: the content of an agent’s knowledge, and the description an agent gives of his knowledge, may include anything individuals do. The content, then, is not what determines whether an action is end-directed. What does is the form that

one's knowledge must take—that is, the logical order displayed, or represented, in practical knowledge. This form of representation is, of course, what Anscombe called 'intention', and what I have been calling 'end-directed'. When an individual's action is end-directed, then, he knows what he is doing, and why doing it is an appropriate step to the realization of one or another aim.⁸ It is because the agent's knowledge conforms to the end-directed structure that his thoughts and actions are unified.

3.3 Vogler On Acts of Expression

Vogler endorses the general view I have just described, but she has a particular way of talking about what the structure of practical knowledge amounts to. For her, the form of description unique to end-directed actions displays a calculative or part-whole order. When an agent is doing something on purpose, we can ask "Why are you x-ing?" and anticipate a certain form of response. Vogler writes:

When Anscombe's hypothetical questioner asks "Why are you A-ing?" or "Why did you A?" or "Why are you going to A?" he forces his interlocutor to look to the calculative form. . . . "In order to B," given in response to the characteristic question "Why?", describes what is taking place. (2002, 130)

So, end-directed actions have a certain calculative or part-whole structure, which is revealed when an agent responds to the question "Why are you x-ing?" Notice, too, that this form of describing what one is doing picks out the means-end or part-whole relation. Again, from Vogler:

What it is for the question "Why are you A-ing?" to have its characteristic [Anscombian] sense . . . is for it to be answerable and made intelligible in terms of how A-ing serves a further end, in terms of the pleasure of A-ing, or in terms of what's fitting or suitable about A-ing under these circumstances given some larger scheme in one's life. (2002, 47)

⁸Anscombe allows for the possibility that an agent can act intelligently "for no particular reason"—that is, for no particular end. In such cases, it may appear that the agent does not know why his action is an appropriate step for realizing one or another aim. In large measure, this appearance is illusory. When pressed, the agent will eventually provide the point of his action as something that is either useful, befitting, or pleasurable to do. Indeed, if there really is *no reason* for the action, it is not an intelligent action. In other words, an agent who cannot provide a reason for his action because he is unaware that he is acting is not acting intelligently. Similarly, if an individual cannot offer a reason for his action that makes its aim something to be desired—that is, if an agent cannot characterize the point of what he is doing as desirable—it is difficult to believe that he has a reason to act in an effort to achieve it. See, for example, Anscombe's "saucer of mud" example in *Intention* (2000, 70-72).

Or, again:

Let A and B be actions of different types. One has reason to A only if one takes it that A-ing is a means to (or part of) attaining (or making it possible to attain) a further end, B-ing, and one wants to B (for no particular reason, or because B-ing is pleasant, useful, or fitting). (2002, 48)

Vogler's view, then, is that when acting on purpose, an agent can describe what she is doing as the means to an end, which she wants to do for no particular reason, because it is pleasant, befitting, or useful. The thought is that since doing anything at all requires taking the means (parts) necessary for realizing the end (whole), this relation must figure in the explanation of action if it is to be characterized as intelligently performed. This structure is revealed in response to a single, simple query: when the question 'Why?' is given application by an agent, the answer displays the form of thought characteristic of end-directed actions, which, of course, reveals the agent's practical knowledge. If that is right, then on Vogler's view, as with Anscombe's, when acting on purpose, an agent knows what he is doing, and why doing it is an appropriate step to the realization of one or another aim. The Anscombian model endorsed by Vogler couples thought to action via a particular form of description. It is when an agent's action and thought share in the calculative form that what an individual does is what the individual thinks he is doing. As Vogler writes: "the calculative form revealed in answer to Anscombe's 'Why?' question belongs to intentional action as such" (2002, 130).

3.3.1 The Challenge of Expressive Acts

Given Vogler's commitment to the Anscombian question 'Why?' as a marker of end-directed action, the possibility of thoughtful action decoupled from the form of description called 'intention' might seem to threaten the Anscombian model. If agents are capable of acting on purpose without being able to articulate what is being done or why it is a step toward their aims, the Anscombian model may, as Austin hints, ignore, overlook, or distort other interesting forms of action. Indeed, Vogler worries that expressive actions—that is, "purposive acts born of mood, sentiment, or temperament which primarily express the agent's state of mind" (2002, 233)—may

threaten the view that all intelligent actions are governed by the calculative form.

She writes:

Expressive action is supposed to pose a problem [for the calculative view] because the reasons don't go deep enough. Expression is supposed to be *essentially* spontaneous. ... [And] action that appears a spontaneous expression of temperament, character, or feeling ... seems to elude the calculative form of description because calculation truly is not in the spirit of the thing. (2002, 233-234)

Vogler worries that expressive actions, because they seem to be done without calculative thought, threaten the link between thought and action that depends on that structure. Her response is to argue that:

Strong emotion can take various primary ends (for instance, to get hold of a photograph or to smash things). The kind of calculation that happens in expressive action with such ends happens when one, say, takes aim in smashing things or uses some means to get hold of a photograph. What matters is that the calculative articulation of events is in place here. If we station our "Why?" questioners at various points along the path of an agent acting expressively, this becomes clear. (2002, 235)

To the objection that the connection between thought and action may be threatened by expressive actions, Vogler responds by insisting that if we *forced* individuals to think about what they were doing, they would provide a calculative articulation of their action. Whether an agent actually thinks before acting is beside the point; the structure of action ensures that if the individual were to think, that thought would take the calculative form.

3.3.2 Art as Expressive Action

Vogler is mistaken, however, to characterize expressive action as *essentially* spontaneous—that is, as essentially removed from thought—and she is mistaken to think that the articulation of expressive actions must conform to the calculative structure. If that is right, thought and action may be connected in other ways, which require other forms of representation—that is, other models of intelligent action—in order to be understood.

To see that expressive actions are not essentially spontaneous, consider that much of what we do from mood, temperament, or sentiment is not merely perfunctory outburst of emotion. The use of art to express oneself, for example, is commonly

held to be more than merely emotional eruption. Indeed, it is typical to regard a variety of types of works of art as displaying thoughtfully controlled expressions of deeply felt emotion. This is evidenced by the fact that individuals perform and hear music as melancholy, joyful, or angry; we also create and see paintings as depressing, lively, or terrifying; the ballet can be moving, calming, or haunting; and so on. To convey mood through art, to express the reality of one's emotion in art, it must be performed well and in a manner that requires thought. If that is right, expressive acts need not be essentially removed from thought. Consequently, they stand as a class of counterexamples to the idea that expressive actions are *essentially* spontaneous.

But if expressive acts are not spontaneous, must the form of thought governing them display the calculative structure? In other words, must the form of thought and action governing artistic expression display the part-whole, or means-end, structure typified by end-directed action? No, not if they can be shown to be done without practical knowledge—that is, not if it can be demonstrated that they resist articulation according to the logical order characteristic of such actions.

To see that they are not governed by that order, consider a few examples. When asked by David Sylvester to explain his artistic process, the American painter Philip Guston responded:

you know, it's terrible to rationalise about painting because you know that, while you're creating it, you can have all sorts of things in your mind consciously that you want to do and that really won't be done . . . yet, when the thing comes off . . . *it arrives at a unity that I never could have predicted and foreseen or planned.* (Sylvester, 2001, 88, emphasis mine)

Similarly, consider Robert Rauschenberg's response to a comparable question from Sylvester:

it happens quite often that I think what the painting needs is a little red right over there and by the time I get the red on the brush and get back to the picture I can't remember where I thought it was to go. But there I am with red and there's the picture and I put it down. And then that's much more interesting for me than sort of building a picture as one might build anything. I prefer the attitude of the picture just evolving *rather than working towards some kind of conclusion.* (2001, 137, emphasis mine)

In another interview, Sylvester asks Willem De Kooning, "So in these paintings what

sort of an idea do you begin with?” Kooning’s response: “I don’t think I set out to do anything” (2001, 54).

And finally, when asked to explain his work, Francis Bacon responds:

I don’t really know how these particular forms come about. I’m not by that suggesting that I’m inspired or gifted. I just don’t know. I look at them—I look at them, probably, from an aesthetic point of view. I know what I want to do, but I don’t know how to do it. And I look at them almost like a stranger, *not knowing how these things have come about and why have these marks that have happened on the canvas evolved into these particular forms*. . . . [Painting] will only catch the mystery of reality if the painter doesn’t know how to do it. *And he’s carried along by his passion and he doesn’t perhaps even know quite what these marks will make*. (Sylvester, 1980, 100–102, emphasis mine)

The number of quotations suggest that it is not unusual for artistic expression to resist articulation according to the calculative form. In fact, each artist explicitly refrains from providing a description of what he does in terms of the calculative structure. For these artists, there are no specifiable ends to which their efforts are directed that would explain what they are doing and, in fact, they seem to lack practical knowledge about what they are doing. Consequently, it would be a mistake to insist that if they *really* thought about what they were doing, their thinking would connect up with their action in virtue of the calculative structure. So much, then, for Vogler’s view that expressive actions *really* must be, at the end of the day, calculative in nature. The thought and action of artists do not seem to be unified by that logical order.

3.4 Befittingly Artistic

But Vogler has something to say about artistic expression as a counterexample to her Anscombian model of action. Indeed, she develops a lengthy example to address complaints like those I have just been raising. Vogler’s discussion is framed by Jack Spicer, a twentieth-century poet who articulates a view of the poetic process similar to the one expressed about painting by Guston, Rauschenberg, Kooning, and Bacon. For Spicer, writing good poetry is being moved along by something seemingly foreign. He writes:

you start seeing whether you can *clear your mind away from the things that are you, the things that you want, and everything else*. Sometimes it’s a twelve-hour struggle to get a ten-line poem, not changing a single

word of it as you're writing, but just as it goes along, trying to distinguish between you and the poem. (1998; 2002, 7 and 120, respectively, emphasis mine)

Writing poetry, Spicer suggests, is not about pursuing what one wants; it is, instead, about ridding oneself of one's wants. That is not to say, however, that one's mind is blank. Indeed, during a conversation with James Reid, Spicer is asked, "do you have any idea . . . where your poem's going to go? I mean, do you have any ideas in your own mind, or any feelings?" Spicer's response: "I try not to." Reid tries to press him on this point by asking him if his mind is a blank. Spicer expands his thought:

No, it isn't [a blank], unfortunately. It's trying to be a blank. And trying to be a blank is utterly different from being a blank. . . . you can't really make your mind a blank. You can't really get to receiving God, which St. Ignatius wanted, or poems, or doing anything. You can't. It's impossible. There's this utter animal spirit which is coming out and saying, well, gee, I can lay this person if I write this line, and all sorts of things like that. It's just impossible to make your mind a blank. (Vogler, 2002; Spicer, 1998, 118-119)

The point, as far as we are concerned, is that on Spicer's view, writing poetry is not a matter of doing things according to the part-whole or means-end structure that seems to govern end-directed actions. When writing poetry, one does not have any idea where things are going. Instead, as with the views about painting expressed by Guston, Kooning, Rauschenberg, and Bacon, Spicer's conception of poetry is one in which the individual frees himself of his ends, and instead, allows the poem to simply "come through" his pen, as though it were being dictated by him with no end in view.

Rather than treating actions with this sort of description as counterexamples to the Anscombian model, Vogler argues that a poet's actions are ultimately governed by that structure.⁹ Let me explain. Although she grants that an artist's *reasons*

⁹There is, I think, genuine ambiguity in Vogler's discussion of the relation between a poet's thought and action. On the one hand, she is clear that she is attempting to frame a type of counterexample to her own calculative view of practical reason. Accordingly, the *reasons* from and for which a poet like Spicer acts are supposed to have no calculative structure. Given that she believes the calculative structure of practical reasons are to be read off the structure of actions, her discussion of the *reasons* from and for which poets like Spicer act implies that the *actions* of poets have no calculative structure either. On this way of understanding her example, a poet's action of writing a poem would be something like a basic action, or a series of basic actions with no internal structure—that is, they would be actions with no parts. On the other hand, however, Vogler talks in several places about a poet's actions being guided by the ends of poetry, which are

for acting—for example, the reasons from and for which Spicer writes poetry—are deeply noncalculative, the poetic process is nevertheless one governed by a “patterning principle” that is ultimately structured by an awareness of ends. Such ends, however, are internal to the patterning principle, which is itself *not* conditional on achieving any other ends. As she writes:

A befitting-style consideration is *radically interminable* if the only ends it generates and serves are internal to the principle in question (for example, if I pursue art for art’s sake alone, then the whole order lent to my life by aesthetic concerns is only intelligible as an aesthetic order pursued for the sake of its aesthetic merit). (2002, 101)¹⁰

In this case, the patterning principles of poetry are what organizes the life of a poet *qua* poet. Vogler writes, “What poets do *qua* poets . . . is done solely from and for the ends of poetry” (2002, 123). On this way of seeing things, the action of writing a poem by a poet is an instance of intelligent action because the poet can give application to Anscombe’s ‘Why?’ question in terms of what it is befitting for a poet to do. For example, asking Spicer, “Why are you *x*-ing?”, where *x*-ing is writing a poem, he can give application to that question by answering, “I am *x*-ing because I am a poet and it is befitting for poets to *x*.” According to Vogler, then, the ordering principle of thought and action—that is, the form of description or ‘intention’—that unifies Spicer’s thought with the act of writing poetry is the ordering principle of his life as a poet. Why are you writing poetry? Because I am a poet, and it befits a poet to write poetry.

internal to the patterning principle befitting the life of a poet. In this sense, the poet’s *actions* have internal structure and are nonbasic actions, which, presumably, means they can be modeled with the Anscombian view. The poet’s *reasons*, however, do not have this structure and cannot be modeled with it. On this interpretation, then, the reasons from and for which agents act do not share the structure of the actions being performed. I do not know which of these two possible interpretations is correct, but whichever it turns out to be, my points will apply, because either way, there is an important class of actions that cannot be modeled with the Anscombian structure endorsed by Vogler.

¹⁰Following Anscombe, Vogler distinguishes three different types of considerations that can serve as the point of action. Another way to say this is that Vogler distinguishes three regions of the good, at which actions can aim. Useful-style considerations give the point of action in terms of what will help an individual achieve a future aim. Pleasure-style considerations provide the point of action in terms of enjoying what is being done in the moment. And, finally, befitting-style considerations give the point of action in terms of its role in patterns that shape one’s life. Radically interminable befitting-style considerations, then, give the point of an action by showing that what is done conforms to a pattern that is, in some sense, self-contained. Such a pattern may shape one’s life and actions, but it does not make use of any other region of the good. The pattern is good in itself.

In response to my worry that artistic expression cannot be articulated in a manner that can be modeled with the Anscombian view, Vogler seems to suggest that poets see their lives, and the actions that make their lives up, as organized by the ends of poetry. Consequently, the Anscombian model, which requires all intelligent action to be done with practical knowledge, gains a foothold on the activities of artists and other agents who cannot articulate their specific ends: although they do not perform their actions with the aim of achieving any further end, they organize their lives and actions according to the patterning principle suited to being an artist, poet, or whatever. It is the ends internal to that patterning principle that govern a poet's action and serve to unify the things a poet does with his or her thoughts.

3.4.1 Against Befitting Explanations

Vogler's argument that the actions of poets, to the extent that they are end-directed, are governed by the patterning principles befitting the life of a poet, warrants consideration. Indeed, I want to suggest that modeling the activity of poets and artists in the way suggested by Vogler—as activity governed by ends internal to the patterning principle befitting a poet's life—really distorts their activities rather than help us to understand them.

So what ends are internal to poetry as a patterning principle to which poets are suited? Consider the following passage from an introductory textbook on poetry.

Poetry enables us to know what it “feels like” to be alive in the world. What does it “feel like,” for instance, to be in love, to hate somebody, to be conscience-stricken, to watch a sunset or stand by a death-bed, to be willing to die for a cause or live in a passionate devotion to some chosen ideal? Only poetry . . . can help us to answer such questions, and help us, thus, to an understanding of ourselves and of our own values. . . . Poetry, it is clear, is not cut off from life, but is basically concerned with life—that is, with the lived fullness of the world. It extends our own limited experience by means of imagination. By imagination, it sharpens our sense of the physical world on the one hand, and on the other, it deepens our sense of the emotional, intellectual, and moral implications of human situations and actions. (Brooks and Warren, 1976, 9)

If we interpret writing poetry as intelligently performed action—that is, as something

a poet does purposefully—its aims include expressing what it feels like to be alive, understanding ourselves and our values, and deepening our sense of the human situation. On Vogler’s view, the life of a poet is patterned after these ends because poets perceive them as internal to the patterning principle befitting the life of a poet—that is, it befits a poet to act like a poet, which means pursuing ends internal to the patterning principle of poetry. In terms of the model she endorses, then, poets *A* because it befits a poet to *A*, where *A*-ing is an end internal to a poet’s life and the agent is or aspires to be a poet (Vogler, 2002, see, for example, 114).

Now the problem with such a view is not so much that it is an impossible model of the actions of poets (or artists), but that it represents the guiding thought of the poet as something that is unlikely to produce poetry (or art). To put that differently, the actions of artists can certainly be modeled in the way proposed by Vogler, but such a model distorts, rather than illuminates, the expressive character of art. Remember that the point of the Anscombian model is to show the unity between what we do and what we think by portraying the shared formal structure of thought and action. According to that model, thought is action when practical knowledge is the cause of what it understands. To produce the unity required by that model, then, the poet must know what he is doing and why doing it is an appropriate step to the realization of one or another aim. In this case, the poet must know what he is doing when writing poetry and understand why doing it is an appropriate step to the realization of the ends internal to the patterning principle that befits a poet’s life. But in the case at hand, it looks as though poets cannot have either component of practical knowledge.

In the first place, it is implausible to suggest that knowing, or having the thought “writing poetry is an appropriate step to the realization of the patterning principle that befits a poet’s life” is what produces poetry, and it is equally implausible to believe that that knowledge can guide action in a way that results in poetry. To be poetic, it is not enough to try to be a poet—by, for example, getting one’s words to rhyme, or getting rhythm into the writing, or organizing the writing into stanzas. The art of poetry is not found in merely writing poems and acting like a poet; rather, the art is had by, apparently, being carried along without really knowing what one is doing. For a poet to write a poem that attains the ends internal to the

patterning principle that befits the life of a poet, the thought that goes into it cannot be controlled by the aim of acting like a poet. If it could, the thought governing the action would be at the fore of the poet's mind, and the poet would not struggle to articulate what he is doing and thinking when writing poetry. Indeed, if merely acting like a poet by writing poems were sufficient to produce *poetry*, writing poems and being a poet would be the same thing. That they are different is betrayed by the fact that many individuals have spent considerable time and effort writing poems without being mistaken for poets.

But that is not all that is wrong with modeling expressive actions with the view endorsed by Vogler. Again, the point of her model is to show that thought and action are unified when one is engaged in intelligent action. But on Vogler's view, the *thought* governing acts of poetry is at a level of control different from the one needed to govern the *act* of writing a poem. Consider that on Vogler's model, the form of thought that governs acts of poetry—that is, the movement of thought and action that produces poetry—is tied to considerations of befittingness. The form of thought, then, would have to be something like “Write a poem (A) because that is what befits the patterning principle of a poet's life (B).” That form of thought, however, is not a plausible reflection of the thought needed to produce poetry. If one's thought and action in writing a poem are to reflect the other's structure—that is, if the Anscombian model is going to get *any* traction on the things poets do—there must be a finer-grained level of control. The thought that produces poetry cannot be that it befits a poet, because that thought does not have the structure needed to, for example, control the movement of the pen. Consequently, if the Anscombian model is going to work, if the actions of poets are going to be describable as intelligent, the poet's thoughts will have to exhibit a finer-grained level of control. A thought like “Do *A* (write these two lines this way) in order to do *B* (express a feeling of grief)” might do the job, but there is no room for that level of control on a befitting-style characterization of such actions. And, anyway, to get Vogler's model to that level of control, she would have to ignore Spicer's description of the poetic process (which, I should emphasize, is something to which Vogler is trying to stay true). She is aware of this fact and, as a result, tries to retain Spicer's description. But by taking this

route, we are left without any explanation of poetic actions: applying the model at the lower level does not capture the phenomena described by poets, and applying it at the higher level cannot account for the control needed to produce the action of writing a poem. What we end up with, then, is either no explanation at all or a distorted view of the poetic process.

3.4.2 A Dilemma

Let me be clear that I am not arguing that the means-end, or calculative, structure *cannot* be read into these activities and used to model them. Certainly artists and poets are doing things, and certainly what they are doing—that is, the process of artistic expression—is composed of parts that together constitute the finished product. Nevertheless, the thought that goes into the activity cannot be modeled with an end-directed form of thought. If it could, its practitioners would not struggle and resist formulating their activities in a manner consonant with that model. Consequently, we should reject the idea that the actions of artists are governed by practical knowledge of the kind espoused by Anscombianism. To insist otherwise is to insist on a distorted representation of that form of action.

This argument presents a dilemma for Vogler (and the Anscombian model more generally): she can either stick to that model and lose the connection between thought and poetry (and other forms of artistic expression), or give up the model and try to locate some other form of thought governing such actions. To do the former would be a mistake because it would suggest that writing poetry is not an intelligent action. If a philosophical account of action forces us to say that artistic expression cannot be intelligent because the thought that goes into it is not of the right form, it seems that we are giving up on the task of explaining action in order to retain a particular philosophical view. This leaves the second horn of the dilemma: to find a different model to represent expressive actions. I will have something to say about that alternative in due course, but before I do, I want to bring in a couple more examples of expressive action—examples that do not depend on interpretations of poetry, painting, or art—that do not fit well with the Anscombian model.

3.4.3 Emotional Expression

To see what I have in mind, consider modes of emotional expression that resist articulation with the calculative form, because they are not governed by ends-in-view. Imagine, for instance, actions that express the grief of losing a loved one. One might go on long, listless walks in the evening, sit at the gravesite and verbalize one's feelings, become somber in one's dress and style, and so on. Such actions seem to be done under the control of the agent, and in that sense, on purpose, but they do not seem to be governed by a form of thought that is directed by one's aims, goals, or ends. The bereaved do not listlessly wander through the streets to get anywhere, or verbalize their feelings in order to communicate, or dress down in order to mourn. Rather, they do these things as acts of emotional expression.¹¹ If that is right, this variety of action seems to be governed by thoughts that give expression to emotion in the same way that an artist expresses his emotion through painting: as non-end-directed actions that express one's feelings. As a result, the unity of thought and action that makes their performance intelligible will not best be modeled by representing the agent as trying to accomplish something.

Indeed, it would be odd to suggest that agents undertake the actions associated with grief in order to achieve one or another type of end. If asked, for example, "Why do you wander through the streets at night?" the bereaved might respond, "Because I feel lost," or because "I'm sad," or because "I do not know what else to do." But the reasons for acting are not governed by the thought that "wandering through the streets at night is befitting to my loss," or the thought that "wandering is pleasant and helps to relieve the sadness," or "it is useful for me to wander aimlessly in order to x ." The wandering is not governed by an end-directed form of thought, and it is

¹¹One might worry that actions like long, listless walks, dressing in a somber manner, verbalizing one's feelings, and so on seem to have the structure typical of the Anscombian model. After all, if an agent does not put one step in front of the other on his long, listless walks, he is not going to go anywhere. Similarly, if an individual does not put his pants on one leg at a time, he is not going to get dressed. But that worry applies to painters, poets, and artists too. I am not denying that we can represent actions like these as actions governed by the calculative form of thought—that is, as actions done with one or another end in view. Sure we can represent things that way. My point is not that all actions cannot be modeled with the Anscombian model; my point, rather, is that there are certain types of action that if modeled in that way leads to a distorted understanding of the relation between the things we think and the things we do.

not done to accomplish anything. Indeed, if the actions of the bereaved were governed by such a form of thought, we could not see them as *expressions* of grief.

Similarly, imagine giving expression to the feeling that comes from learning that the one with whom you have fallen in love reciprocates that feeling. What does one do to express how one feels? Write a poem, perhaps. Or, more sensationally, prance whimsically through the streets, swinging from lamp posts, and generally making a lighthearted fool of oneself (I'm thinking of the kind of behavior typified by, for example, Gene Kelly in the 1952 movie "*Singin' in the Rain.*") Such actions are not governed by thoughts directed toward ends. Imagine, for example, asking Gene Kelly's character as he is dancing through the streets, "What are you doing?" The answer, I believe, is unlikely to be anything other than something that appeals directly to the emotion he feels. When an individual is overwhelmed by love, he does not go dancing through the streets for any reason other than to express emotion. And such actions need not be seen as merely thoughtless outbursts of emotion. Although the things we do on such occasions do not have the coordination and sophistication of Gene Kelly's performance, they still require thought. Such thought, I am suggesting, is different than the end-directed thought governing other forms of intelligent action. It is thought governed by the emotion it expresses. If that is right, then, as with expression through art and poetry, thoughtfully acting to express how one feels should generally resist the form of description typical of Anscombian views.

Furthermore, even if expressive actions like those of the bereaved were governed by the end-directed structure, they would require the agent to have absurd beliefs about his ends. As Rosalind Hursthouse argues in "Arational Actions," there are no "nonabsurd candidates for appropriate beliefs to ascribe to agents performing [expressive] actions" (1991, 60).¹² Consider again a bereaved individual listlessly wandering through the streets of his town because he is saddened by the death of his lover. What could such an action be aimed at trying to achieve? It is not as

¹²Hursthouse uses "arational actions" in place of my "expressive actions." Other than the terminological difference, there is very little to distinguish them. Indeed, Hursthouse argues that acts of emotional expression (her "arational actions") stand as a class of counterexamples to the thought that all intentional actions can be described with means-end relations. The argument of this paragraph relies heavily on the strength of that argument.

though his wandering is going to bring the loved one back or make the sadness go away. And it is not as though such wandering is done because it is befitting, useful, or pleasurable. One does not listlessly drift through the streets because it befits one's grief.

Indeed, even if such wandering is befitting of the emotion, it is not as though a bereaved agent wanders in order for his action to be befitting of his grief. If he did, we would see his action as, in some sense, disingenuous, as something done in a playacted manner rather than as a genuine expression of emotion. An agent who acted in a manner characteristic of the bereaved because it befit his grief would be displaced from his emotion; it would be the action of someone who acted because he thought that such actions were what someone who feels what he is supposed to feel would do. It is a confusion, then, to think that such expressive acts are governed by ends of any sort. In fact, interpreting the actions of an agent in the grips of a particular emotion as governed by ends befitting the emotion forces us to attribute to them modes of expression that are alien to anyone who has genuinely acted from the relevant emotion. Adopting Vogler's position to model expressive actions, then, forces us to insist on representing the actions of individuals as though they were alien. Put differently, if expressive actions must be articulated according to the Anscombian model of action, and if they often aim at ends removed from the emotions they express, then for an agent to act expressively requires that she take steps to realize ends that she cannot understand.

But to the extent that such actions are intelligent, the Anscombian model endorsed by Vogler is forced to represent them in this way. As a result, on that model, such actions must be articulable according to the end-directed form. When asked, "Why are you wandering aimlessly through the streets?" or "Why are you singing in the rain?" the answer must be something on the order of "I'm *A*-ing in order to *B*, and I want to *B* because it is fitting, useful, or pleasant."¹³ But articulating expressive

¹³As I have mentioned, Vogler's view also makes room for an agent to respond to the Anscombian question "Why are you *A*-ing?" by saying, "No particular reason." In this way, the bereaved might respond by saying, "I'm wandering aimlessly through the streets for no particular reason." But even if this response is available to the bereaved, the point of the Anscombian model (and Vogler's model along with it) is to show that thoughts and actions share a logical structure when the latter are done in an end-directed manner. If that is right, saying "I am *A*-ing for no particular reason" is only

actions in this way is contrived. The natural response, and the response that explains the actions of individuals in the grips of emotion, does not appeal to ends. Instead, the response that explains the action appeals to the emotion being expressed. “Why are wandering through the streets?” “Because I feel lost at my loss”; “Why are you singing in the rain?” “Because I’m in love!” In such cases, the agent’s appeal to emotion is what explains the action, not any presumed ends directing its performance.

As a result, there must be actions that express states of emotion that are not governed by thoughts ordered by the structure found in the Anscombian model of action explanation. We might want to frame them that way, but to do so is to insist on a theoretical position that seems at odds with what really lies behind such actions.

3.5 Toward an Alternative

Since I am about to suggest a different way to think about expressive actions, it will be useful to pause and recapitulate. I began by outlining the Anscombian model of action explanation. According to it, intelligent actions are performed with practical knowledge. In other words, when an agent knows what he is doing, and is able to locate and articulate what he is doing in terms of a part-whole or means-end relation, his thought and his actions are unified by a form of description that shows his action to be intentional. Vogler characterizes this form of description as calculative, and I have just finished arguing that her model can only represent expressive actions by distorting them. This leads to a dilemma: she can either ignore the distortions and stick to the Anscombian model; or give up the distortions by finding a different way to model expressive actions. In this section, I want to gesture at a way to think about the latter half of that dilemma.

a legitimate answer—that is, one that shows the action to be done intentionally—if, after further thought, the individual could articulate her action according to the calculative form of thought governing it. The idea, then, is to suggest that even when the ends governing one’s thought and action are not immediately before one’s mind, they are still there governing the action. On this way of understanding things, the action of the bereaved is intentional even if he resists articulating some particular end as the one at which his action is aimed because it is still governed by some end or other. This suggestion, however, is precisely the one I am arguing against. When an agent resists providing a calculative articulation of his or her actions, it is a mistake to insist that that structure is there governing the action anyway. There are other options for modeling the connection between thought and action.

3.5.1 Ordering Emotion

In the case of artistic expression, it is clear that the actions of the artist are more than merely spontaneous outbursts of emotion. Indeed, I want to suggest that such actions are, instead, ways of ordering and clarifying emotion. I will use that suggestion in a moment to argue that, in general, acts of emotional expression—like singing in the rain when overcome with joy or wandering listlessly through the streets because of one's grief—are constrained by an order of thought similar to that governing modes of artistic expression. Before I get to that, I need to explain what it means to say that artistic expression is a way of ordering emotion.

Consider a few different examples of theorists trying to explain art as an act of emotional expression. John Hospers writes:

[Artistic creativity] is [a process of] clarifying [the feeling] to himself, [an individual] cannot before expressing it state what he is going to express; therefore he cannot *calculate* in advance what effects he wants to produce and then proceed to produce them. (1954-1955, emphasis in original)

In a similar vein, R.G. Collingwood writes:

Until a man has expressed his emotion, he does not yet know what emotion it is. The act of expressing it [through art] is therefore an exploration of his own emotions. He is trying to find out what these emotions are. (1958, 111)

John Dewey embraces a similar view of artistic expression when, in *Art as Experience*, he writes:

While there is no expression, unless there is urge from within outwards, the welling up must be clarified and ordered by taking into itself the values of prior experiences before it can be an act of expression. And these values are not called into play save through objects of the environment that offer resistance to the direct discharge of emotion and impulse. Emotional discharge is a necessary but not sufficient condition of expression. (1934, 61)

Taken together, these authors seem to be suggesting that acts of emotional expression are governed by modes of resistance that clarify and order the emotions as they are expressed. As the thought goes, then, artists do not allow emotion to simply propel them headlong into spontaneous emotional outbursts, but instead, clarify and order their emotions by forms of resistance that stem, if Dewey is right, from values of prior experience. It is by frustrating the uncontrolled discharge of emotion (through

poetry, painting, sculpting, and so on) that the artist effectively expresses what he feels.

A similar attitude toward artistic expression is conveyed by Francis Sparshott when he writes:

the arts of dance use the movements in which human feelings about actions and situations find expression as the basis of art forms in which these and like movements are clarified, made more visible, and reduced in order.
(1997, 120)

Again, the idea seems to be that dance is based on forms of emotional expression that are, in some sense, purified by that style of artistic expression. My point, of course, is not about particular styles of artistic expression, but is instead about what it means to give expression to emotion, which is in contrast to having an emotional outburst. In each of the quotes above, it seems that the form of expression is governed by modes of resistance that clarify and order feeling and emotion, not by particular ends or aims. If we can understand the nature of these forms of resistance, we can get a better sense of how to model the link between thought and expressive action.

3.5.2 Emotion and Thought

So what do these forms of resistance look like? Is there a discernible order to the way in which emotions are expressed? My own sense is that the order is to be found in the values that frame emotions themselves. Let me try to explain what I mean by that. Ronald de Sousa has argued that emotions influence our perception of a realm of value. An individual's apprehension of the world is influenced by his feelings of anger, sadness, grief, joy and so on. This perceptual shift is often expressed metaphorically: angry individuals see red, everything is sunny and bright to those in the grips of joy, and to those who are sad and depressed, the world seems grey and drab. Such metaphorical expressions represent an acknowledged shift in the significance we attribute to the world when in the clutches of emotion. Assume, then, that that is right, that the significance we attribute to things shifts with variations in emotion. The structure of those shifts is what I mean by saying that the order to be found in the thoughts that govern emotional expression is in the values that frame emotions themselves.

Let me try to explain that idea more fully. The fact that we use metaphors to express the perceptual shift in our worldview when influenced by emotions shows that we do not understand emotions very well. Nevertheless, such shifts shape our actions, and I believe it is such shifts that artists are trying to capture by giving expression to their emotions through art—that is, by various modes of artistic expression, individuals show the perceptual shift effected by emotion. It is as though the struggle of the artist is to reveal how things are when the significance they attribute to the world shifts due to emotion. To accomplish that requires, to use Spicer's imagery, to let oneself go and let the feeling flow through one's pen. Or, as Bacon puts it, to be carried along by one's passion without quite knowing what marks the paint will make on the canvas. Only in this way, both Bacon and Spicer seem to think, will the artist be able to capture the mystery of reality. That mystery, I am suggesting, is the perceptual shift in value that originates in emotion, which is only captured if, as Dewey points out, one resists the direct discharge of emotion by clarifying and ordering it through action.

That is all very conjectural. Nevertheless, it provides an avenue for making sense of the idea that individuals, and artists in particular, act intelligently even though they do not seem to have the type of practical knowledge required by Anscombian views. Since the perceptual shift typified by emotion is understood, at best, only metaphorically, actions that give expression to that shift can only be understood metaphorically. This explains why artists have such a difficult time articulating what it is that they are doing when they are doing art. The suggestion, then, is that artists act without practical knowledge because they do not know, until they have expressed it, what it is that they are giving expression to. The thought governing the actions of an artist is the emotion it expresses, but since the artist only understands that emotion metaphorically and through action, he cannot articulate the structure of the thought that he is giving expression to.

Of course, one thoroughly committed to the Anscombian view is likely to respond that even if all of this is right, the artist is still trying to do something: he is still trying to give expression to something that he only understands metaphorically. In other words, he is trying in some sense to *A* in order to *B* (because it is useful,

befitting, or pleasurable). But that is a mistake. The thoughts that control the expression of emotion through art are not governed by that structure. They are governed by the form of the emotions themselves, and it is their reality—that is, the reality of emotions that give shape to our thought—that artists are expressing when acting from emotion. To express one’s emotions, an individual has to let go and act without practical knowledge while, at the same time, clarifying and ordering his feelings through action.

If I am right, expressive actions stand as a class of counterexamples to the Anscombian thought that intelligent actions have a single structure with which they can be modeled. Acts of emotional expression call for a different model of action, one that does not rely on an agent’s practical knowledge. When we give expression to our emotions, we do not know what we are doing or why we are doing it. Our thought and action is governed by the emotions themselves. But that does not mean that such actions are unintentional, unintelligent, or merely spontaneous outbursts of inner mental states. Wandering through the streets after the loss of a loved one may be an action governed by the emotion itself, and may seem to be mindless wandering, but more often than not, the action is an expression of an agent coming to terms with thoughts that have taken a different form, and such “coming to terms” is done by clarifying, ordering, and shaping them through action.

3.6 Conclusion

There are limits to what any single representation can portray. The Anscombian model reaches those limits when it attempts to portray expressive actions with a model that depends on practical knowledge. As we have seen, it is frequently the case that we act intelligently without knowing what it is we are trying to accomplish with our actions. Consequently, modeling such actions with a form of representation built around the calculative relation will have difficulty portraying them. At best, modeling expressive actions with the calculative relation leads to a distorted explanation of the things we do from emotion; at worst, understanding expressive actions with that structure undermines attempts to explain how emotional thought is tied to the things we do. I have tried to show what those distortions look like and what is lost when

we seek to explain expressive actions using the Anscombian model.

Central to what I have said is the idea that portrayals of what it means to act for reasons—that is, representations of rational agency—are models. As such, they will be suited to particular purposes. Anscombian models work well for explaining end-oriented actions, but they struggle to make sense of expressive actions. For that purpose, we need a different model of action. And that fact is an indication that what we are trying to understand—that is, our own sense of being rational creatures—is a complex phenomenon whose nuances cannot be captured in a single, simple model. And this is a lesson that brings us back to Austin. As he noted:

It must be remembered that there is no necessity whatsoever that the various models . . . should all fit together neatly as parts into a single, total model or scheme of, for instance, the doing of actions. It is possible, and indeed highly likely, that our assortment of models will include some, or many, that are overlapping, conflicting, or more generally simply *disparate*. (151 Austin, 1961, emphasis in original)

Following this thought, I have tried to show that expressive actions are peculiar forms of human action. If that is right, a fuller picture of intelligent action, and the reasons that explain the things we do, will require more, and better, models.

CHAPTER 4

ACQUIRING KNOW HOW

4.1 Introduction

Many of the things people know how to do never had to be learned. We never had to learn how to push things, how to suckle, how to scratch an itch, how to see, hear, and breathe, or how to think basic thoughts. But a number of other things that we know how to do, we had to learn how to do. Riding a bicycle, driving a car, skiing, mountain climbing, playing chess or poker, writing philosophy papers, playing the piano, and so on are all skills that had to be learned in order to know how to do them. What explains the difference? What did we have to learn to acquire capacities of the latter sort that we did not have to learn to know how to do things of the former sort? In this chapter, I am going to argue that the difference is explained by the things we must do to coordinate thought with action.

This explanation, however, raises a more fundamental question: is learning to coordinate thought with action something we do on purpose, as itself a type of action? If so, what type of action is it? There is not a viable philosophical answer to these questions, or so I will argue. In fact, to clear the way for answering them, the majority of the chapter will be spent arguing against views that explain intelligent action by, in one way or another, presupposing the unity of thought and action. For views of this sort, the things we do and the things we think are so tightly interwoven that acquiring information about how to do something entails knowing how to do it. More simply, acquiring knowledge-how is had by acquiring knowledge-that. But such a view is problematic: to learn how to ski, ride a bicycle, or drive a car, it is not enough to learn the information pertaining to the actions that go into effectively skiing, cycling, or driving a car. In addition, agents must learn to coordinate the things they do with the thoughts that govern them in a manner that results in the performance of intelligent action. More generally, learning how to F requires learning both the

content of the thoughts that go into *F*-ing as well as how to coordinate those thoughts with the activities required for effectively *F*-ing. Learning the latter, I will argue, is not something that can be explained by appealing to thought alone. Indeed, I think that explaining how agents learn to coordinate thought with action requires thinking anew about action itself.

The motivating problem of the chapter, then, is to explain the acquisition of a certain type of knowledge-how by explaining the things we do to coordinate thought with action. To get started, it will be useful to clear a bit of terrain. To that end, I am going to use ideas from Gilbert Ryle's *The Concept of Mind* to distinguish two different ways of thinking about how to explain action. In Section 4.3 of this chapter, I argue against one of these options, which has recently been defended by Jason Stanley and Timothy Williamson. According to them, learning-how is learning a fact.¹ I argue that such a view fails to explain what we learn when we learn how to do things. With their view out of the way, I return to Ryle to get a clearer sense of the type of knowledge-how we are trying to explain. In Section 4.4, then, I distinguish between capacities agents simply possess and those that must be learned, and suggest that, given the former, it might be possible to develop the latter through thought alone. In Section 4.5, I argue against one way of making the case for that idea. Recently, Sebastian Rödl has suggested that agents who possess a basic power for practical thought can learn to do things they have never before done by merely reflecting on what they already know how to do.² I believe this view is mistaken, and I will try to show why. In Section 4.6, I gesture at an alternative framework for explaining the acquisition of knowledge-how. The shape of that idea depends on an analogue of induction to explain how thought and action come to be coordinated. This alternative, I suggest, shows promise for thinking about how to design models

¹The meaning of words like 'fact' and 'proposition' are notoriously controversial. I want to remain agnostic about controversies and opt instead for a fairly pliant posture. Accordingly, I assume facts are true propositions, and propositions are the objects believed or asserted when an agent believes or asserts, for example, that snow is white.

²Although this feature of his view is not obvious, I believe he is committed to it. As a result, I will spend a good portion of Section 4.5 showing that, in fact, Rödl is committed to the idea that we can learn new skills by thought alone.

of action that can explain the things we do to acquire knowledge-how.

4.2 Practical Intellectualism

Ryle allowed that there might be two kinds of knowledge that serve to explain the performance of intelligent action. On the one hand, the performance of intelligent actions might be explained by the apprehension of true propositions—that is, by an agent’s knowledge-that. On the other hand, their performance might be explained by an agent’s understanding of how to do things—that is, by knowledge-how. The difference between these two types of knowledge, however, is obscured by the fact that Ryle uses the term ‘knowledge’ for both. It suggests that both forms of knowledge are, in some sense, informational, and interpreted this way, they both seem to refer to the substance of thought, or, in other words, to the type of information contained in thought.

For reasons that will become clear, I think this way of interpreting Ryle is a mistake. Nevertheless, I want to flag the idea that intelligent action can be explained by appealing solely to the content of thought and give it a label. Strong Practical Intellectualism (SPI) holds that knowing how to do something is knowing that something is the case; it is knowing facts of one variety or another. According to SPI, the factual information encoded in propositions is necessary and sufficient for explaining the thoughts that go into intelligently doing things. This view has recently been defended by Jason Stanley and Timothy Williamson, and crucially depends on showing that information about oneself—one’s capacities, dispositions, and abilities—can be encoded in propositional knowledge and account for everything that is known when an agent knows how to do something.

But there is another way to interpret Ryle’s distinction between knowledge-that and knowledge-how. According to this alternative, knowledge-that is characterized as any and all of the informational content of thought. (Information about the world, about one’s capacities, one’s dispositions, one’s abilities, how to do things, and so on are all instances of knowledge-that: in other words, anything that can be stated with the phrase “I know that”) In contrast, knowledge-how is not, in any sense, informational. It is, rather, a capacity for action of one sort or another. (The capacity

of insects to walk, for example, is knowledge-how on this sort of interpretation. So too, the capacity of rational agents for thought is, on this interpretation, knowledge-how.) Now, although putting things this way leaves residual ambiguity in Ryle's discussion—after all, if the capacity for thought is knowledge-how, it is far from clear whether that capacity can be exercised without informational content—as with SPI, I want to flag the idea that explaining intelligent action requires more than what can be encoded in the content of thought and to give it a label. Weak Practical Intellectualism (WPI) holds that propositional knowledge is necessary for explaining intelligent action, but, in contrast to SPI, it is not sufficient to explain what is known when an agent knows how to do something. In addition to the information encoded in thought, WPI requires that those thoughts be governed by a specific capacity, or, to put that differently, a specific way of thinking, in order to produce intelligent action. For my purposes, I am going to associate WPI with Anscombian models of intentional action.

The difference between SPI and WPI, then, turns on the properties of thought needed to explain intelligent action.³ Proponents of SPI believe that the content of thought is all that is needed to explain intelligent action; in contrast, proponents of WPI argue that to explain intelligent action, one must appeal to both the content of thought and to a *specific* way of entertaining that content.

From both strands of intellectualism, it seems to follow that agents can learn how to do things by simply acquiring propositional information. SPI is explicit in this commitment, arguing that since knowledge-how is knowledge-that, learning-how is nothing more than learning-that.⁴ WPI, however, is more subtle. According to it,

³I am going to use the phrase 'intelligent action' where the alternative 'intentional action' might do just as well. There are two reasons for this terminological choice. First, the term 'intention' has come to mean many different things in recent philosophical literature, and I want to avoid confusing associations. Second, many of the ideas I will be engaged with in this chapter are rooted in Gilbert Ryle's *The Concept of Mind*. In that book, and in the second chapter of that book in particular, he is interested in understanding what an agent must know in order to perform *intelligent* actions. I will stick to Ryle's turn of phrase except in Section 4.5.1, where, for exegetical purposes, I use 'intentional action'.

⁴This supposed entailment is stated explicitly by Stanley in the preface of his recent book. He writes: "The thesis of this book is that knowing how to do something is the same as knowing a fact. It follows that learning how to do something is learning a fact" (Stanley, 2011a, vii).

learning truths about how to F results in knowing how to F when the information encoded in thought is entertained in a particular way, or according to a capacity for thought of a particular form. On both types of intellectualism, then, learning how to F can be accomplished by acquiring the information encoded in propositions about F -ing. I will return to WPI in Section 4.4, but for now, I want to show that SPI cannot explain what agents learn when they learn how to do things.

4.3 Intellectualizing Knowledge-How

Although discussing knowledge-how inevitably requires addressing Ryle's work, I am going to postpone doing so until we have worked through SPI. As I have mentioned, Strong Practical Intellectualists believe that the thoughts that go into intelligent action can be explained by appealing solely to their propositional content. The view is developed by Stanley and Williamson in "Knowing How" (2001), and has since been augmented and defended by Stanley in his monograph *Know How* (2011a).⁵

Both iterations of the view support SPI by defending a particular schema. As Stanley writes:

To defend the view that knowing how to do something is a kind of knowing that something is the case, it is sufficient to defend the validity of the following schema: *For every s and F , s knows how to F iff for some way w of F -ing, s knows that w is a way to F .* If this schema is valid, then knowing how to do something is a species of knowing that something is the case. (Stanley, 2011a, 71, emphasis mine)⁶

As I read this schema, it is necessary and sufficient for knowing how to F that an agent knows at least one way in which he could successfully F .

But what does it mean to know a way of doing something such that we would ascribe to an agent knowledge-how? According to proponents of SPI, the meaning of "John knows how to F " is determined by analyzing the informational content of the thoughts tied to F -ing. This content includes, first, the information known when an

⁵Stanley has also responded to criticisms and developed his ideas in "Knowing (How)" (2011b). In addition, he defends the idea that knowing-how is a species of knowing-that in a co-authored piece; see (Pavese and Stanley, 2011).

⁶For the earlier iteration, see (Stanley and Williamson, 2001, 430).

agent knows the answer to a question embedded in an ascription of knowledge-how. For example, if John knows how to ski, he knows of at least one way to ski such that doing things in that way could result in his successful action. The agent, then, relates to the information encoded in the proposition in virtue of knowing an answer to the embedded question *how does one ski?* Second, the information known by John is first-personal information. When John knows how to ski, he knows that it is he himself who knows that there is a way to ski such that doing things in that way could result in his successfully skiing. So when John knows how to ski, the information contained in the thought that goes into skiing is first-personal thought about what he himself knows how to do. Third, the infinitival phrase “to ski” has a certain modality. In the case at hand, for John to know how to ski is for John to know of all the many ways one *could* ski that there is at least one way in which he himself *could* ski successfully. In other words, of the various ways in which individuals effectively ski, John need not know them all; rather, his knowing how to ski only depends on his knowing at least one way in which he himself could ski. So in sum, on this account, the information encoded in an individual’s knowledge-how is represented by the semantical information contained in propositions understood by the agent. When this information is possessed by a particular agent—that is, when an agent stands in the right relation to the information encoded in these propositions—we ascribe to him knowledge-how.

I do not want to resist the general idea that agents who know how to do things typically possess some or all of the information specified by Stanley and Williamson. It seems right to say that ascribing knowledge-how to an individual generally depends on that individual grasping information about how to do whatever it is we say he or she knows how to do. Furthermore, I do not see any reason to resist the idea that that information, whatever it turns out to be, can be encoded in propositions. Yet, even granting this, and even granting that the relevant information may be indefinitely rich and complex, there seems to be an obvious objection to make to SPI. In particular, it seems that the informational content of thought is not itself sufficient to account for the capacities involved with knowing how to do things. Even if it is true that possessing propositional knowledge is necessary for knowing-how to *F*,

that information is not sufficient for F -ing. This is most straightforwardly seen by considering the manner in which knowledge-how is acquired.

4.3.1 Acquisition

It is very easy to acquire the informational content of propositions. After all, if that information is what is common to the thought ‘snow is white’ and the sentence ‘snow is white’, there is little reason to think that it cannot be had by simply reading the sentence. Information is easy to come by. So if knowledge-how is propositional (as proponents of SPI would have it), agents should be able to learn how to do things simply by acquiring the relevant information.

But such a view of learning how is problematic because translating information about ways of doing things into a capacity to do them seems to require much more than simply obtaining information. Consider John’s sister, Hannah, who is trying to learn how to ski. She pesters John incessantly about ways of skiing (which, recall, are the facts known when an agent knows how to F). She asks about techniques for skiing in different snow conditions, the postures needed for effectively navigating one’s way downhill, methods for stopping, for controlling oneself, and so on. In addition, she has watched videos of John skiing while he has pointed out all the various nuances used to navigate runs in different conditions, she has watched countless instructional videos on line, she has read books, and she has spent hours memorizing information about the positions and techniques required for skiing. In short, she has done everything except ski in order to learn how to ski. Yet all this information is insufficient for acquiring the capacity to ski. Her first time out, she confidently clicks into her bindings, struggles toward the chairlift, then, when attempting to load, promptly falls over. She does the same thing getting off the lift and throughout the remainder of her first day. But why? After all, we are assuming that she has *all* the relevant information. Why, then, does not Hannah know how to ski? Because the information encoded in propositions about skiing is not sufficient to explain what agents learn when they learn how to ski. Information alone is not suited to the practical nature of acquiring knowledge-how.⁷

⁷Notice the similarity between this argument and Frank Jackson’s well-known argument in “Epiphenomenal Qualia” (1982). Jackson argues that Mary, an individual who has only experienced the world within a black and white room, but who has all physical information about colors, will

We learn to do things, as Aristotle pointed out, by doing them, not by acquiring information about ways of doing them.

4.3.2 Modes of Thought

There are two closely related strategies used by proponents of SPI to get out of this problem. The first, mentioned by Stanley and Williamson in their original piece, addresses the issue by arguing that the informational content of propositions tied to knowledge-how must be thought of according to a “particular mode of presentation”—alternatively, a “particular way of thinking” or “mode of thought”—in order to result in successful action. The second strategy is deployed by Stanley alone when, in his more developed monograph, he argues that differences in the modal parameters of propositional information explains why an agent like Hannah does not have knowledge-how, despite the fact that she has what is *ex hypothesi* all factual information. Let me consider each in turn.

Stanley and Williamson argue that the content of an agent’s thought must find expression according to a particular mode of presentation—or a particular way of thinking—in order to be thought that demonstrates knowledge-how. The idea is to tie the information encoded in an individual’s thoughts to different ways of thinking in order to explain how the same information can give rise to different effects. The most straightforward example of this idea relies on the distinction between demonstrative and first-personal ways of thinking. Consider the following passage from Stanley and Williamson:

Suppose that John is looking in a mirror, which he mistakenly believes to be a window. Seeing a man whose pants are on fire, and not recognizing that man as himself, John forms the belief that that man is on fire. Intuitively, however, John does not believe that his own pants are on fire. That is, relative to the *invisaged* context, (26) is true and (27) is false:

- (26) John believes that that man has burning pants.
- (27) John believes that he himself has burning pants.

learn something new about the world when she is let out. In particular, she will learn what it is like to experience color. In virtue of that fact, physicalism—the view that all information is physical information—is false. The argument of this paragraph is similar: if an individual who has all factual information about skiing must learn something else in order to ski, the view that knowing how to do something is knowing a collection of facts is false.

Given that ‘that man’ refers to John, however, the complement clauses of (26) and (27) express the same proposition, namely the singular proposition containing John. To distinguish between (26) and (27), contemporary advocates of Russellian propositions appeal to different modes of presentation under which that proposition is entertained. In the envisaged context, (26) is associated with a demonstrative mode of presentation (or guise) of the relevant proposition, whereas (27) is associated with a first-personal mode of presentation of that very same proposition. (2001, 428)

To explain how the information encoded in propositions explains knowing how to F , Stanley and Williamson make a similar move. Just as identical information presented in either a demonstrative or first-personal way of thinking leads to different results, so too, thinking of that information according to a practical mode of thought has different effects as well.⁸

Consider, for example, the knowledge possessed by Hannah about ways of skiing. Presumably, entertaining the information pertaining to skiing according to a demonstrative mode of thought will not result in her successfully skiing, because that mode of thought is not tied to the dispositions and abilities necessary for successfully skiing. Thinking demonstratively about some way, w , that it is a way to ski does not result in action typical of knowing how to ski, because this way of thinking is not tied to the skills, dispositions, and abilities required for skiing. In contrast, however, entertaining that information under a practical mode of presentation does lead to successfully skiing, because that mode of thought is tied to the requisite dispositions and abilities. The idea is to argue that thoughts with the same propositional content can have different effects depending on how they are entertained. Since the demonstrative mode is not tied to practical capacities, but a practical mode is, entertaining the information encoded in propositions about how to ski in the latter case, but not in

⁸It is not entirely clear what, exactly, a “practical mode of thought” amounts to for Stanley and Williamson. In particular, they never make clear how the distinction between, say, demonstrative, first-personal, and “practical” ways of thinking can be made by appealing to what seem to be the different *actions* resulting from the different modes of thought. If demonstrative and first-personal modes of thought are distinguished by their effects, are they not both, in some sense, “practical”? I am not sure how Stanley and Williamson would answer that question, but, regardless, my concern is only with practical modes of thought, which they seem to think are unique in that they are “related in complex ways to dispositional states” (2001, 430).

Later in the chapter, when I look at WPI and Anscombian views of action, I will try to spell out in more detail what a “practical mode of thought” could plausibly amount to. Of course, those details will be independent of the view endorsed by Stanley and Williamson.

the former, results in successful action. Stanley and Williamson's response to the challenge that information is not sufficient for knowing how to ski, then, is to argue that Hannah may have all the knowledge necessary for claiming that she knows how to ski, but be unable to ski, if the content of her thought is not entertained in the relevant mode of presentation.

In his monograph, Stanley acknowledges that practical modes of thought may seem unappealing, and, although he maintains his commitment to the idea, provides another avenue of defense to explain why an agent might seem to have all the relevant information about how to F while still not knowing how to F . The alternative strategy is to argue that the information encoded in propositions pertaining to knowing how to do things is sensitive to modal parameters tied to an agent's physical abilities. In this way, what an agent knows when she knows how to F is sensitive to what she is capable of doing. Consequently, an agent can *seem* to have all the relevant information about how to F while not knowing how to F , if the information built into the relevant propositions is not transparently tied to the right modal parameters. Once the latter are taken into consideration, however, the difficulty is resolved, or so Stanley argues.

To see the problem, and Stanley's solution to it, more clearly, consider two propositions (altered from Stanley (2011a, 126)).

15a) Hannah knows that the way John skis is a way in which she could ski.

15b) Hannah knows how to ski

Ignoring for a moment what Stanley calls the modal parameters tied to Hannah's physical abilities, the information contained in either 15a or 15b should be sufficient to account for Hannah's knowing how to ski. After all, if knowing how to F is knowing that for some way w of F -ing, w is a way to F , then on either 15a or 15b, Hannah should know how to ski. But, of course, that seems false: Hannah may know that the way John skis is a way to ski without knowing how to ski herself. There is an asymmetry between 15a and 15b and the worry is that Stanley's view cannot account for it.

Stanley meets the worry, however, by appealing to the modal parameters tied to Hannah's physical abilities. As he writes:

the difference in meaning [between 15a and 15b] ... is due to the different modal parameters governing the uses of the modal constructions in these sentences. ... In (15a), the modal parameter is one that takes the world of evaluation, and yields a set of propositions that characterize Hannah's physical state after training for some time ... In contrast, the natural modal parameter for the envisaged utterance of (15b) is one that takes the world of evaluation, and yields a set of propositions that characterize Hannah's physical state at the moment. That is why the two utterances express different propositions—because the modals in the two sentences are interpreted via distinct modal parameters. (2011a, 126)

Stanley seems to think that the informational content of propositions tied to knowledge-how may include what one is physically capable of doing, and that by building this information into the thoughts that go into knowing how to do something, an agent's knowledge-how can be represented wholly by appealing to what is encoded in propositions.

Notice that both the earlier appeal to particular modes of expression and the later appeal to modal parameters rely on the idea that the information encoded in propositions is sensitive to an agent's dispositions and abilities. This is how proponents of SPI retain their commitment to the idea that the propositional content of thought is sufficient to explain what an individual knows when he knows how to F . According to the first response, the information encoded in propositions is sensitive to its mode of presentation. Only if the information is thought of in a particular, practical way, will it result in actions that demonstrate knowledge-how. The second response is similar: the content of the information encoded in knowledge-how is sensitive to contexts of interpretation, which take into account the physical dispositions and abilities of agents in particular contexts. As a result, a proper interpretation of the informational content of thought depends on understanding what the agent can do—that is, on the modal parameters framing the agent's know-how. The shared idea, then, is expressed in arguments aimed to show that an agent knows how to F only if she knows at least one way of F -ing such that that way is a way to F , and she has the relevant dispositions and abilities required for performing the action in question.

4.3.3 Impending Regress

It is difficult to see how these responses are adequate to the problem faced by proponents of SPI. The problem is that explaining knowledge-how using only what can be encoded in the propositional content of thought presupposes an agent's capacity to know how to entertain that information in a manner that leads to the relevant actions. Such a presupposition cannot be maintained if one is trying to articulate what it means to know how to do something. This should be familiar to anyone acquainted with the work of Ryle. What Ryle noticed was that to explain intelligent action, one cannot appeal solely to propositional knowledge, because, in addition to it, intelligent action requires agents to possess certain dispositions and abilities, which cannot be explained propositionally. That may sound similar to what proponents of SPI have to say. After all, they argue that in addition to the informational content of propositions, one must know how to entertain it according to a practical mode of thought, or know how to execute it with the dispositions and abilities one possesses according to the relevant modal parameters. But, in contrast to Ryle, SPI tries to show that knowing how to entertain the informational content of thought according to a practical mode of thought can be accommodated with information about facts, or knowledge-that. This solution, however, leads to an explanatory problem foreseen by Ryle.⁹

If knowledge-how is a species of knowledge-that, the content of the thoughts that go into knowing how to think according to a practical mode of thought must be specifiable with further propositional information. But adopting this strategy results in a failure to explain how it is possible to entertain that information in the first place. To see why, suppose that it is true that the thoughts that go into knowledge-how are fully specifiable with propositional information. If that is right, then the thoughts that go into entertaining the information encoded in propositions with a practical mode of thought must be explained by specifying the information encoded in propositions (this, of course, is just to repeat the commitments of SPI). But to think of the

⁹The problem was identified long before Ryle by Lewis Carroll in "What the Tortoise Said to Achilles" (1895). Carroll demonstrates that the reason to accept the legitimacy of an inference cannot be a further premise to the inference.

information encoded in knowing how to think with a practical mode of thought requires that the information be thought according to a particular mode of thought. Consequently, an agent must know how to think of the information encoded in knowing how to think according to a practical mode of thought using a practical mode of thought. But since thinking according to that mode of thought is something we know how to do, that information can be fully specified with the information encoded in propositions as well. Consequently, to entertain the propositional information specifying the mode of thought used to entertain it, we must use the mode of thought we are aiming to specify. The problem should be clear: if thinking is an intelligent action, the thoughts that go into it cannot be fully specified by appealing to the information encoded in propositions without leading to a vicious regress. Adopting this route, as proponents of SPI do, fails to explain how it is that agents think the thoughts required for doing the things they do, because it presupposes, and does not explain, the capacity of agents to know how to do things.

Ryle makes a nearly identical point when he writes:

The consideration of propositions is itself an operation the execution of which can be more or less intelligent, less or more stupid. But if, for any operation to be intelligently executed, a prior theoretical operation had first to be performed and performed intelligently, it would be a logical impossibility for anyone ever to break into the circle. (1949, 30)

To get thought up and running, it seems, there must be something more than the propositional content of those thoughts; there must be some basic dispositions or abilities tied to an agent's capacity for thought, which are used to execute the information encoded in propositions. This, of course, is Ryle's solution. From the paragraph before the one just cited, he writes:

It is therefore possible for people intelligently to perform some sorts of operations when they are not yet able to consider any propositions enjoining how they should be performed. Some intelligent performances are not controlled by any anterior acknowledgments of the principles applied in them. (1949, 30)

Ryle characterized the dispositions, or abilities, to perform some type of intelligent operations as a distinct type of knowledge.¹⁰ Indeed, for Ryle, knowledge-how is

¹⁰I think Ryle's appropriation of the term 'knowledge' to describe these dispositional states may

dispositional knowledge, which is intended to be a type of knowledge that can make sense of being and acting intelligently without depending on one's grasp of propositional truths. To explain what is known when agents perform intelligent actions, Ryle offers two different types of knowledge: one is dispositional, the other propositional.

Proponents of SPI seem forced to a similar conclusion, yet they maintain that the content of thought can do all the explanatory work. But that is a mistake, because it cannot avoid the regress problem Ryle was worried about. Furthermore, trying to avoid that problem by appealing to practical modes of thought, or to the modal parameters governing knowledge-how statements, is not a solution either. If proponents of SPI appeal to either—that is, to either practical modes of thought or modal parameters—as a disposition or capacity necessary for knowledge-how, and if neither are reducible to propositional information, then by their own lights, the schema that proponents of SPI aim to defend is false: it is not the case that *s* knows how to *F* iff for some way *w* of *F*-ing, *s* knows that *w* is a way to *F*. If an agent does not possess the capacity or ability to think about the information encoded in propositions in a manner productive of action, he or she may have all the relevant information and still not know how to *F*.

4.3.4 Learning to Act

So SPI cannot, without regress, explain the knowledge-how possessed by an agent when he or she knows how to think practically—that is, SPI cannot explain the basic capacity of rational agents to entertain propositional information in a manner that leads to action. But that is not SPI's only problem. In addition, there are a pair of problems pertaining to the acquisition of knowledge-how.

The first problem is closely related to the regress problem. To explain what agents learn when they acquire knowledge-how, SPI must assume something that needs explaining. In particular, the view must assume that, given an agent's capacity to think according to a practical mode of thought, providing new propositional

have been a mistake. Be that as it may, his point that we need to understand how to do things and that this understanding cannot be encoded propositionally was not. In other words, even if it is not accurate to describe certain basic capacities for action as states of knowledge, Ryle was certainly correct to distinguish the idea as something required for intelligent action.

information is sufficient to account for the acquisition of knowledge-how. But that assumption cannot be justified unless we can explain how the power to entertain information according to a mode of thought produces such knowledge. In other words, appealing to the information encoded in propositions cannot explain how individuals acquire knowledge-how unless we already understand how intelligent action results from “practical modes of presentation,” “practical ways of thinking,” or the “modal parameters” required for interpreting practical thought. Without explaining what these phrases mean and how they lead to intelligent action, SPI cannot explain the acquisition of knowledge-how by appealing to propositional information. The capacity to entertain information according to a practical mode of thought cannot be assumed when attempting to explain the things we learn how to do, because, for all that has been said, entertaining information in that way might itself be something we must learn how to do. If it is, we cannot explain what we learn when we acquire knowledge-how by appealing to it.

In addition to explanatory regresses, SPI runs into yet another difficulty pertaining to the acquisition of knowledge-how. In particular, the view entails that knowledge-how can be acquired by inference. To see this, suppose SPI is true. That is, suppose it is true that knowing how to do something is just knowing that something is the case; it is knowing information that can be encoded in the content of thought. Given such an assumption, an agent should be able to derive knowledge-how by inference.

Suppose, for example, that Jones knows how to do two different things, each of which can be done without an understanding of how to do the other. For example, Jones knows how to ski (S) and Jones knows how to perform backflips on a trampoline (B). Given the assumption that SPI is true (and this means that we are assuming that SPI does not have the regress problem), the thoughts involved with his doing S and the thoughts involved with Jones’s doing B should be fully specifiable with the information encoded in the relevant propositions. Furthermore, since we are assuming that the information encoded in thought can *also* account for its mode of presentation, that information should be sufficient to account for what is learned when Jones learns a new capacity, say, SB-ing. In particular, Jones should be able to derive from the information pertaining to S and from the information pertaining to B, the information

pertaining to SB-ing. From two independent skills, an agent should be able to derive a third, composite skill since, by hypothesis, he has all the information.¹¹ In other words, given the truth of SPI (and ignoring the regress problem), an individual should be able to simply conjoin the propositional information responsible for his knowing how, and, in virtue of that conjunction, know how to do something else.

Given SPI, this follows from fairly uncontroversial assumptions about propositions. After all, it is standardly accepted that the content of propositions is subject to basic principles of logic and can be manipulated by those principles to derive new propositions. If I know that “this is snow,” that “this is white,” and that “this is cold,” then it is uncontroversial that the knowledge that “this snow is white and cold” can be derived by thinking about the information encoded in the propositions. But such powers of inference do not seem to apply to cases of knowing how: if I know how to entertain the information pertaining to skiing with a practical mode of thought, and I know how to entertain the information pertaining to doing backflips according to the same mode of thought, and if I know how to ski and I know how to do backflips, I still will not be able to derive the knowledge needed to do backflips while skiing simply by reflecting on the information that I already know and can entertain with a practical mode of thought. This fact suggests that there is something more to the combined capacity than the information encoded in the thoughts pertaining to either skiing or backflipping, and this is true no matter how rich and complex we make that information.

4.4 Basic Powers and Intelligent Actions

Given the points of the previous section, it looks like we need more than what can be accounted for with propositional information to explain the acquisition of knowledge-how. In addition, we need (minimally) to explain what it means to

¹¹One might worry that there must be additional information required for the performance of the composite skill. For example, it seems plausible to suggest that in addition to the information pertaining to the individual skills, one must also know information pertaining to how they are integrated in the performance of a single activity. Maybe that is right, but if it is, we should wonder how that information is acquired. I am suggesting that it cannot be acquired by obtaining information merely. Indeed, much of the remainder of the chapter is dedicated to explaining how the information pertaining to integrated skills may (and may not) be acquired.

entertain thought according to a practical mode of presentation. This, I think, is a point recognized by Ryle, and in this section, I want to return briefly to Ryle's account. Doing so will help clarify the issue before us and frame the discussion of WPI, which is still to come.

Ryle uses the term 'knowledge-how' to refer to two seemingly distinct things. In response to what I am calling his regress argument, Ryle argues that there must be basic dispositions, skills, or abilities for intelligent action that do not rely on the "anterior acknowledgment" of the information governing their performance. I am going to call these *basic powers*.¹² Candidates include things that agents know how to do that were never learned, yet seem to be instances of intelligent action. For example, the power to push objects, to suckle, to scratch an itch, and to think and reason (in some minimal sense) seem to be candidates for such basic powers. They are things an agent either knows how to do or does not, and if he does not, there is no information that is going to help him learn. If I do not know how to make an inference, for example, there is no information that could show me how.¹³

In contrast to basic powers, there are skills and abilities that agents must acquire. I will call these further abilities, following Ryle, *intelligent capacities* or *skills*. Contenders include any skilled activity: making an omelette, mountain climbing, skiing, riding a bicycle, driving a car, and so on. Ryle characterizes intelligent capacities by contrasting them with habits. He writes:

When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having to mind what he is doing. He does not exercise care, vigilance, or criticism. ... But a mountaineer walking over ice-covered rocks in a high wind in the dark ... thinks what he is doing, he is ready for emergencies, he economises in effort, he makes tests and experiments; in short he walks with some degree of skill and judgment. (1949, 42)

Intelligent capacities, or skills, require agents to think about, and to be aware of, what they are doing while they are doing it. Yet it is clear that the performance of a

¹²It does not matter for my purpose whether Ryle actually held the view I am using him to develop. As a result, I am going to simply assume that the two senses of knowledge-how I am about to distinguish were recognized by Ryle, but not adequately separated because he thought they were, in some sense, too tightly interwoven.

¹³This, of course, is one lesson of Carroll's "What the Tortoise Said to Achilles." See footnote 9.

skill or intelligent capacity is not an expression of a basic power. Instead, they seem to be the products of learning. If that is right, there is a difference between basic powers and intelligent capacities, though Ryle never adequately distinguishes them.¹⁴

Given that they are different, consider again the chapter's motivating question: how can we explain the acquisition of knowledge-how? As a question about basic powers, it is a nonstarter. As Ryle suggests, there must be some intelligent operations that are performed without an antecedent understanding of propositional truths. This fact follows from his regress argument. If that is right, our motivating question should be understood as one aimed at explaining the acquisition of intelligent capacities or skills. So how can we explain their acquisition? As we have already seen, appealing to the procurement of information encoded in propositions is not going to do the job. If it could, it would be possible to learn to ski, mountain climb, or ride a bicycle by acquiring information or inferring the skills from information pertaining to what we already know how to do. How, then, do we acquire intelligent capacities?

There is some potential for explaining their acquisition by assuming certain basic powers. For example, one might assume that rational agents have the power to entertain the content of thought according to a practical mode of presentation. In this way, an agent that possesses that power can acquire intelligent capacities by simply acquiring the information contained in propositions and entertaining it according to the appropriate mode of thought. But, as we have seen, unless we specify what that power is and how it works, we do not really have an explanation of how acquiring information can result in the acquisition of intelligent capacities. This is one of the problems that sunk SPI.

So what could a basic power for action be? Is there a basic power of thought that, when combined with propositional information, causes intelligent action? Anscombians believe that they have an answer to this question because they have tried to provide a substantive account of practical thought—that is, they have tried to specify the form of thought that causes action. For Anscombians, when thought is entertained

¹⁴In the sense I am using the terms, both basic powers and intelligent capacities or skills can produce intelligent actions. For example, I can use the basic power to push things to push open a door in order to go shopping. Similarly, if I possess the skill to mountain climb, I can use that skill, and the knowledge-how associated with it, to maneuver up a rock-face.

according to a specific pattern, it results in action. Despite their more robust view of action explanation, however, I do not believe that they can use it to explain the acquisition of intelligent capacities or skills.

4.5 Intellectualized Action

As I have argued, the problem with SPI is that it tries to explain the know how needed to perform intelligent actions by appealing solely to the content of thought. But the view collapses in the face of what it cannot explain. One way to try to salvage the view is by moving to a type of practical intellectualism that does not rely solely on thought's content to explain knowing how to do something. To that end, I want to sketch a version of WPI, which, recall, is the idea that thought must be governed by a particular formal structure—that is, the content of thought must be entertained in a particular manner—in order to produce intelligent actions. As we will see, the idea aims to specify what it means to entertain thought's content according to a basic power for practical thought, one whereby thinking thoughts in that way causes action.

The work of Elizabeth Anscombe and her contemporary followers is, I think, representative of the type of Weak Practical Intellectualism (WPI) that I want to use as a target. For Anscombians, the possibility of intelligent action depends on the unity of one's thought with one's action. When an individual does something on purpose for a reason, thought and action unfold together according to a shared structure. A commitment to this type of unity, however, seems to imply that new skills and intelligent capacities can be acquired through thought alone. In other words, if thought and action share a tight enough connection, one should be able to figure out how to perform skilled actions by acquiring the information about how to perform them. (It is not obvious that this is an implication of the view. The argument that it is will be provided in a moment.) But accepting this seeming implication of WPI is a mistake. Indeed, what I aim to show is that the view of intelligent action put forward by proponents of WPI presupposes without explanation the capacity of agents to acquire new skills and intelligent capacities. To make that argument, I want to show that the *basic power* for practical thought put forward by WPI cannot explain what we learn when we acquire new skills or intelligent capacities, because

it cannot account for the actions performed to acquire them. I begin by delineating the contours of the type of Weak Practical Intellectualism characteristic of Anscombe and her recent followers.

4.5.1 Anscombian Action

For Anscombe, the distinguishing feature of intentional action is not the content of one's thought, but rather, its form or structure.¹⁵ When acting intentionally—that is, in a manner such that one's actions are governed by one's thoughts—the content of thought must take a particular form; it is this formal feature of thought that ties it to action. This is not, of course, to suggest that proponents of WPI think that the substance of thought is irrelevant to an explanation of intelligent action. After all, it is presupposed that agents who act intentionally have knowledge of what they are doing. As Anscombe notes, “it is the agent's knowledge of what he is doing that gives the descriptions under which what is going on is the execution of an intention” (2000, 87). When acting intentionally, one knows what one is doing and can articulate that knowledge in a manner that demonstrates an understanding of facts. Nevertheless, it is not the informational content of an agent's thought that serves to distinguish intentional action; rather, it is the form or structure of those thoughts. Again from Anscombe:

the term ‘intentional’ has reference to a *form* of description of events. What is essential to this form is displayed by the results of our enquiries into the question ‘Why?’ Events are typically described in the form when ‘in order to’ or ‘because’ (in one sense) is attached to their descriptions. (2000, 84-85, emphasis in original)

The idea is that whatever the content of an agent's thought turns out to be, it will provide a description of an intentional action only if it is constrained by a certain formal structure. In particular, it is in virtue of the progressive (or means-end) form

¹⁵Although I am generally using the phrase ‘intelligent action’ instead of ‘intentional action’ throughout this essay, I take exception to that policy in the next couple of paragraphs.

My characterization of Anscombe comes primarily from her *Intention*. That being said, many of the ideas found in *Intention* are, to my mind, more clearly articulated in “Practical Inference” (Anscombe, 1995). In that piece, Anscombe argues for the idea that the form of thought is important for distinguishing intentional actions from other things rational agents do.

of thought,¹⁶ and a rational agent's ability to locate what he is doing in terms of that structure, that an action is demarcated as intentional. If the progress of thought was not coordinated with the progress of action—that is, if one's thought could not take the means-end or part-whole form typical of actions—it would not be possible for what an agent is thinking and what he is doing to be tied together in a self-conscious event. If that is right, saying what one is doing when one is doing something intentionally presupposes that one's thought conforms to the formal structure internal to actions.

Recently, a small, but influential, cadre of theorists has endorsed these Anscombian ideas. They argue that to explain the connection between thought and action, the structure of one must share the structure of the other. Only in this way can thought and action form the sort of causal unity typified by rational agency.¹⁷ I am going to call this idea—the idea that intelligent action depends on the formal unity of thought and action—the *unity thesis*. One implication of this thesis, which presumes an extraordinarily tight connection between the things we think and the things we do, is that we can acquire the knowledge-how typical of skills and intelligent capacities through thought alone. Accepting that implication, however, is a mistake. We cannot acquire the skill to mountain climb, ski, or ride a bike by reflection merely, because the form of thought required to unify what one thinks and what one does, even when combined with propositional information about how to do things, cannot explain how we come to coordinate thought and action in the first place. Before I can argue for that, however, I need to sketch the ideas behind the unity thesis, and show that it

¹⁶Although Anscombe does not herself adopt this terminology, it will prove apt as we proceed. The phrasing is taken from recent work by philosophers influenced by Anscombe. Michael Thompson, for example, writes: “the use of the progressive in the articulation of ordinary event-consciousness seems somehow to span the present, reaching into the future” (2008, 126). And Sebastian Rödl seems to support a similar view when he writes: “A progressive thought looks forward in the sense that it designates *a certain end as proper to what it represents*” (2007, 30, emphasis in original). The idea espoused by both authors is that the progressive form of thought shows the unity of thought and action when agents act intelligently.

¹⁷The idea that there is a certain type of causal unity between thought and action may sound puzzling. The idea, however, is not that bizarre. As Thompson writes: “The nature of intentional action, or the kind of being-subject-of-an-event that characterizes a rational agent and a person, resides in the peculiar “synthesis” that unites the various parts and phases of something like house building, for example, mixing mortar, laying bricks, hammering nails, etc.” (2008, 91). The question, then, is what makes all these different actions the single, unified action of an agent building a house? And the response is, I take it, that it is the agent's thought in doing them.

entails a commitment to the view that individuals can acquire intelligent capacities just by thinking.

4.5.2 Anscombe Updated

Anscombe and her recent followers are, I think, distinguishable by their commitment to one or another version of the unity thesis. The different versions, however, are idiosyncratic enough to be unwieldy if discussed together as a single view. Consequently, I am going to focus almost exclusively on a single iteration of that thesis. To this end, I will be concerned with the work of Sebastian Rödl. There are, however, two further reasons for focusing on him: first, his commitment to the unity thesis and what it entails is straightforward; and second, the view that the unity thesis commits one to the idea that acquiring knowledge-how can be accomplished by thinking alone is most easily shown by looking at his work. Nevertheless, to the extent that Anscombe and her sympathizers accept the relevant aspects of Rödl's conception of the unity thesis, the points I make will apply to them as well. By way of segue, let me begin by saying a word or two about Michael Thompson's view.¹⁸

Thompson argues that the structure of thought “can figure in the order of things equally as grounded and as ground, as rationalized and as (non-finally) rationalizing” (Thompson, 2008, 90). The idea is tied to the thought that in the most basic sense, the concept of ‘intentional action’—that is, the idea that agents are capable of performing intelligent actions—is explained by an agent's self-conscious understanding of his place in a process he causes (Thompson, 2008, 132).¹⁹ In this sense, an agent's thought and action are unified in an “etiological nexus”²⁰ of which the agent is self-consciously aware. This nexus of thought and action must share a specific structure,

¹⁸That Thompson's views should serve as a segue to Rödl's is suggested by Rödl when he writes in the preface to *Self-Consciousness*: “I do not think there is a thought in this book that was not at some time or other the topic of conversations I have had with Michael Thompson. If a thought in this book is of value, I shall not know that it is not his” (2007, xi).

¹⁹As with Anscombe's view, the idea is not to dismiss the content of thought in explaining the relation between thought and action, but rather, to show that thought must be constrained by particular formal parameters in order to be tied to intentional action.

²⁰Thompson goes out of his way to remain agnostic about explicitly causal questions. See, for example, footnote 3 in part 2 of *Life and Action* (2008, 86).

which Thompson demonstrates to be the progressive or imperfect form of event-consciousness (Thompson, 2008, 125-126). Without the capacity to entertain the information contained in thought according to the progressive structure, an agent's thought and action would be, at best, only coincidentally connected, and such an agent would not qualify as a rational agent.

Rödl agrees with Thompson that the coordinated unity of thought and action requires the content of the former to be constrained by the progressive form of event-consciousness. Rödl, however, is more clearly focused on the peculiarly causal aspect of this unity. He writes: "if a movement rests on thought, then the unity of its phases, which constitutes it as a movement, must rest on thought" (Rödl, 2007, 31). Only in this way can what an agent thinks and what he does be tied together. If that is right, the unity thesis depends on there being a very tight connection between the things we think and the things we do when we are doing them intelligently. Indeed, the connection is so tight that Rödl explicitly adopts causal terminology to mark it. He writes:

action explanations do not cite further mental causes, but represent a different configuration of thought, will, and action. When we explain that someone is doing *A* because she wants to do *B*, and add that she thinks that doing *A* is a means of doing *B*, we do not give a further cause; rather, we specify *the kind of causality*. We give the sense of the question "Why?" that we answer. If the explanation is true, then the subject's thought constitutes the causal nexus. The causality of the will is thought. (2007, 50)

This passage entails that practical thought *is* intelligent action and intelligent action *is* practical thought.

That said, the unity of thought and action in rational agency is not solely dependent on the formal structure of event-consciousness. In addition to the shared structure, the performance of intelligent action depends on the informational content of thought. This is just to point out Rödl's commitment to WPI. Indeed, this commitment is evident in Rödl's discussion of the elements of practical reasoning. For him, practical thought is constituted by a practical inference—that is, an inference that is, by its nature, the cause of action. Such inferences depend on two premises. The first is "a desire as Kant defines it: a representation through which its subject is the cause of the existence of its object. The second . . . is a statement of independent

fact; it purports to be (in the good case it is) speculative knowledge” (Rödl, 2011, 219). These two premises, when combined through a basic power to act for reasons, are the cause of intelligent action. The overall picture, then, is one according to which the performance of intelligent action depends on a desire to do something, which is steered by an agent’s speculative understanding of the means required to satisfy it. The desire of the first premise and the knowledge content of the second premise serve to generate movements that reflect the progressive form internal to actions.

An example will prove useful. Suppose an individual wants an omelette and knows how to make them—that is, he desires an omelette in the Kantian sense and grasps the means-end relations necessary for omelette making. Given his speculative understanding of how to make omelettes and his desire to have one, the agent may, as a matter of practical inference, perform the actions that cause the existence of an omelette. In such a case, the agent’s thought about omelette making and the actions involved in making an omelette are tied together by the progressive form they share. Furthermore, we say that the agent possesses the skill, or know how, to make omelettes, because his practical inference is an expression of that skill.

That is all well and good, but the question this chapter is addressing is not “What is practical inference?” but, rather, “how is it that agents acquire intelligent capacities or skills?” That is, how do agents acquire the skill to make omelettes (or to ski, mountain climb, bicycle ride, and so on) such that, given a desire for an omelette (or to ski, mountain climb, ride a bike, and so on), he can perform the intelligent actions that express that skill? In terms of the details of Rödl’s account, the question is, “how does an agent come to grasp the knowledge contained in the second premise of a practical inference such that that knowledge, combined with a desire, effects an intelligent action?”

Calling the knowledge of the second premise “speculative,” as Rödl does (Rödl, 2011, 219), seems to imply that it is grasped theoretically, as propositional information about the means necessary to realize one’s ends. If that is right, if the information contained in the second premise of a practical inference is merely propositional knowledge, then one should be able to acquire intelligent capacities through thought alone. After all, given a desire for F and speculative knowledge about how to do F , there

seems to be nothing to stop a practical inference from being effected. But if that is right, then given the relevant desire, an agent should be able to execute the fulfillment of that desire without ever having done what he has set out to do—that is, the agent should be able to perform the actions characteristic of a skill as a matter of practical inference. Ultimately, I think Rödl is committed to that view, but attributing it to him straightaway is too simplistic and overlooks important nuances of his account. Let me explain.

Rödl insists that “the will, being the source of practical reasoning, is the formal cause of the second premise of practical reasoning” (Rödl, 2011, 224). What this means is that the speculative knowledge contained in the second premise of a practical inference is arranged, or structured, by an individual’s capacity as an agent. To put that differently, “knowledge of means depends on the will for its form; the will depends on knowledge of means for its matter; the matter is inseparable from its form” (Rödl, 2011, 224). The idea seems to be that an individual’s understanding of the progressive structure of action is determined by the will. If that is true, it would seem that individuals grasp the knowledge contained in the second premise by exercising their will—that is, agents acquire knowledge-how by doing things.

But on Rödl’s view, that is only half true. To see why, recall the distinction between basic powers and intelligent capacities. Basic powers are things that agents know how to do without ever learning how to do them. In contrast, intelligent capacities are things agents know how to do that have had to be acquired through learning. And our question has been, “How do agents acquire the knowledge-how typified by intelligent capacities or skills?” Rödl seems to have two different ways of answering that question, which also turn on the difference between basic powers and intelligent capacities.

On the one hand, Rödl agrees with Ryle that there are basic powers possessed by rational agents as such. He calls them “elementary powers of movement” and ties them to general knowledge-how. So, for instance, if, as a rational agent, I possess the elementary power to push things, I thereby know how to push things. From this general knowledge-how, I can derive more specific knowledge-how. For example, given that I know how to push things generally, I can derive the specific knowledge-how

needed to push *this* door open. In similar fashion, given that an agent already knows how to perform an intelligent capacity like, for example, skiing, the knowledge-how required for specific acts of skiing can be derived from that more general knowledge. According to his conception of practical inference, then, a desire to ski, combined with general knowledge-how about skiing, may result in the intelligent action of skiing. An individual's thought is action, because, on occasions like this, an intelligent capacity (knowing how to ski) is used to derive specific knowledge-how (knowledge about how to ski at this moment), which, when combined with a desire to ski, effects a practical inference that is the performance of skiing as an intelligent action.

But of course, we do not have an elementary power to ski; it has to be acquired. So the question is, "What other way is there to acquire the knowledge-how typified by skills?" Rödl has an answer. In addition to deriving it from general knowledge-how, it can be calculated from other things an agent knows how to do. So, if I have speculative knowledge of the fact that I can do *Z* by doing *W*, *X*, and *Y*, and I know how to do *W*, *X*, and *Y*, I can calculate the knowledge-how required to do *Z*. In this way, I can infer or calculate *Z* from things I already know how to do. Again, according to Rödl's conception of practical inference, a desire to ski, combined with the calculative knowledge-how contained in the second premise of a practical inference, may be used in practical thought to effect the action of skiing. On occasions like this, thought is action because the specific knowledge-how (knowledge about how to ski at this moment) can be calculated from what the agent already knows how to do (the parts that go into skiing).

The half-truth in Rödl's view, then, is that the knowledge-how contained in the second premise of a practical inference is acquired on some occasions by *deriving* it from what agents already know how to do. In particular, when the knowledge-how required to perform a specific action is derived from more general knowledge-how about performing some skill, or intelligent capacity, it is acquired from previous exercises of the will. If I have already exercised my will such that I know how to ski, I can use that general knowledge-how to perform an intelligent action by practical inference. The mistake in the view, however, is in thinking that the knowledge contained in the second premise of a practical inference can be calculated or inferred

by thinking about things I already know how to do. Indeed, given this calculative method for inferring the knowledge-how contained in the second premise, it follows on Rödl's view that, if an agent knows how to do the parts constitutive of a skill or intelligent capacity, the agent can acquire the relevant skill or capacity by thinking merely. After all, to calculate a new skill, *Z*, from what I already know how to do, *W*, *X*, and *Y*, there is nothing else to appeal to other than thought alone, since I make that calculation without ever having done *Z*. His commitment to that view is, I believe, betrayed by Rödl on a couple of occasions. For example, he writes:

The power to reason about what to do *is* a power to do things. For, in the fundamental case, thinking that such-and-such is to be done because ... is the causality of an action explanation that one *is doing it* because ... Since practical thinking is, fundamentally, acting, the power of practical thought is a power to act. (Rödl, 2007, 60, emphasis in original)

In a footnote to this passage, he takes things even further by interpreting Aristotle and Kant as supporting the view. He writes:

One manifestation of [the power of practical thought as a power to act] ... is that deliberation about how to do something terminates in things one can do. Another ... is that recognizing that one must do something is recognizing that one has the power to do it. (2007, 60)

These quotes, I think, betray an implication of the unity thesis. For anyone committed to that thesis, thought is action when the premises of a practical inference are satisfied. But an agent can satisfy the premises of a practical inference without actually knowing how to do what he has set out to do. This fact is a consequence of the idea that agents can calculate the content of the second premise of a practical inference by reflecting on things they already know how to do, and from that knowledge-how infer a new skill or intelligent capacity. If an agent knows how to do the parts of an action, he can calculate how to do the whole just by thinking about the relation of the parts to the whole. It follows that agents can learn to perform intelligent actions by reflection merely, even if they have never before performed the particular action. But that is a mistake. Knowing how to perform the individual parts of an action is not sufficient for performing the action as a whole. If that is right, then one cannot calculate knowledge-how to *Z* from knowledge-how to *W*, *X*, and *Y*, where the latter are the constitutive means to the former.

4.5.3 Learning to Act

To see this, consider an agent with full knowledge of the content of the second premise of a practical inference, and that that knowledge pertains to the performance of two distinct intelligent actions, say, skiing and doing backflips. Given an agent in full possession of such information, is it possible to explain the acquisition of a third, composite intelligent capacity by appealing to calculation? I think the answer is no, but there are two ways one might support the alternative view. On the one hand, it might be that an agent in full possession of the information required for performing two distinct intelligent actions can simply reflect on his speculative knowledge and infer the composite skill. Of course, I have argued in Section 4.3 that we cannot make such inferences. Alternatively, then, it may be that an agent can derive the composite skill by using the relevant information plus a basic power for progressive (means-end) thought. In this way, propositional information can, given a power for practical inference of the sort specified by Rödl—that is, a disposition to entertain the content of thought according to a specific formal pattern—explain the acquisition of knowledge-how. This, of course, is to suggest that WPI is the right way to think about acquiring knowledge-how. This alternative, however, must presuppose without explanation the intelligent capacities it aims to explain. If knowledge-how is required for practical inference, and we try to use practical inference to explain the acquisition of *intelligent capacities*, we fall into an explanatory circle.

There are, I think, two types of examples that show why WPI cannot explain the acquisition of new intelligent capacities by appealing to practical inference, but before turning to the examples, let me provide a more general argument. When an individual is learning how to do things, there are many things she can know. She can know information about what it takes to *F*, she can know information about why doing *A* is a means to doing *F*, she can know what she is capable of doing, what she is disposed to do, what she is able to do, and she can even know how to entertain all of that information according to the means-end, or progressive, form of thought. That last point is, of course, a concession to the idea that means-end thought may very well be a basic power in the sense espoused by Ryle. Yet even knowing all this information, and even possessing the power of practical thought, the agent may not know how to

coordinate the performance of the particular actions necessary for accomplishing what she has set out to do. The thought, then, turns on the idea that when one is learning a new skill, there is a gap between the general capacity for progressive, or means-end, thought, the information encoded in thoughts about how to do things (even granting indefinitely rich and complex propositional information), and the execution of that thought in a manner that produces the relevant action. The reason for this is that the general capacity to think practically cannot itself serve to coordinate an agent's actions with the thoughts that produce them—that is, thoughts about what to do in order to F are not coordinated with the bodily movements required to F by reflection merely. To acquire the know how that is displayed by the coordinated unity of thought and action typical of practical inference—that is, to fill in the gap between the general capacity for progressive thought and the particular actions required to effectively do something—we must first act without knowing what we are doing.

Two quick examples will demonstrate this point. The first is meant to show that the *basic power* for progressive thought cannot yield intelligent capacities or skills; the second is meant to show that the unity of thought and action required to perform two independent intelligent capacities cannot be used to derive the unity of thought and action required for a third, composite skill. Suppose Smith has speculative knowledge of how to ski groomed runs and he is attempting to learn how to ski powder. In other words, he possesses knowledge-how about skiing in a somewhat restricted sense, and he is trying to expand his capacities. Moreover, suppose he has a basic power for progressive thought—that is, he can think about the information pertaining to skiing in terms of the progressive form. He can think, for example, that in the conditions in which he finds himself, skiing effectively requires his weight to be (roughly) centered over his bindings, that his knees should be bent so that his shins are pressing on the front of his boots, that his arms should be extended forward, and so on. These thoughts are what he does. We are supposing, then, that Smith has all of the relevant information and that he can entertain it according to the structure relevant to practical thought, which causes his action. Yet, unless he has, through a multitude of attempts, coordinated the connection between the particular actions required for skiing and the thought that governs their

successful execution in powder—that is, unless Smith has forged a connection between his restricted knowledge-how and the activity he is trying to perform—he will not be able to ski in powder. The reason for this is that merely thinking on the information contained in thought with the means-end, or progressive, form is not enough for the things one thinks to be coordinated with the movements required for action. One cannot merely think of things according to a certain formal relation and, in virtue of that thought, execute the movements necessary for performing the action. One's thought and the actions required for executing a skill must be coordinated through a process of learning how to do things, and that process is not one of reflection or calculation merely. This, I think, shows that it is false to suggest that “deliberation about how to do something terminates in things one can do” (2007, 60).

This claim is further supported by considering the fact that agents cannot derive new skills from already possessed skills. I made this point earlier. Even if an individual has forged the relevant connection between thought and action for two distinct intelligent actions, he cannot use them to derive a third, composite skill. But if the general capacity for progressive thought were sufficient to coordinate the things we think with the movements necessary for action, we should be able to infer, or calculate, composite skills from already possessed skills. Why? Because if the general capacity for progressive thought could serve to coordinate thought with the movements necessary for action, there would be nothing else to learn. We would not have to learn anew how to coordinate our movements with the thoughts that cause the performance of the composite skill or intelligent capacity. But, of course, that seems to be a stretch. To acquire the skill to do backflips while skiing, it is not enough to know how to ski and to know how to do backflips. Sure, knowing how to do those things independently may help to acquire the new skill, but it is not sufficient. Agents cannot learn composite skills by deriving the necessary unity of thought and action from what they already know how to do, even if the things they already know how to do are constitutive components of the composite skill. Instead, agents must establish new connections between their thoughts and the movements required for performing actions of the more sophisticated type. And that cannot be done through thought alone. Rather, it must, as Aristotle suggests, be done by doing.

4.5.4 Coordinated Movement

To coordinate the things we do in performing an action with the structure and content of the thoughts required for their execution, agents must first do things without knowing what they are doing. Such action, it seems to me, is a precondition for learning to coordinate thought with the actions that cause the performance of intelligent action, and unless that precondition is satisfied, an agent cannot learn to move his body in a way that characterizes the performance of skills. But if that is right, some of what we do on purpose will not be governed by the progressive form of thought. After all, if I must ski in order to learn how to ski, it cannot be the case that I already know how to perform the action, nor can it be that I already grasp the connection between the actions necessary for effectively executing the skill and my thoughts about skiing. This fact indicates that there is a class of actions involved with acquiring new intelligent capacities that WPI—and Anscombian models of action more generally—is not suited to explain.

Let me emphasize these points. If the arguments I have been developing are right, we are faced with an odd result. To acquire new intelligent capacities, and the knowledge-how associated with them, we must do things on purpose without knowing what we are doing. Furthermore, these actions cannot be represented, or modeled, while accepting intellectualist assumptions about action. It looks like there is a class of intelligent actions, which are tied to acquiring knowledge-how, that theories of action explanation have ignored. How, then, should we represent such actions? In the next section, I want to very briefly gesture at what I think is a first step towards filling this explanatory hole.

4.6 Modeling Explorative Actions

I have argued that neither SPI nor WPI can explain the things we do to acquire knowledge-how. Their failure to do so is a result of their commitment to intellectualist assumptions. According to both views, the information encoded in knowledge-how is required for intelligent action, and, in one way or another, it seems to follow that the acquisition of intelligent capacities can be had by reflection alone. But the inference is a mistake, which is betrayed by the fact that intellectualist models of action cannot

account for the kind of learned coordination between thought and action required for learning new skills. Since the actions involved with acquiring such skills depend on individuals doing things without knowing what they are doing, intellectualism of either variety cannot explain them as intelligent actions. If that is right, giving an account of the actions involved with learning how to do things demands an account of action explanation that does not assume practical intellectualism of either variety.

In the face of this argument, intellectualists might respond by simply suggesting that the things individuals do to acquire knowledge-how are irrelevant to an account of intelligent action. After all, if they are done without knowing what we are doing, why bother trying to understand them? They look like ungoverned actions. Indeed, if such actions are tantamount to behavioral thrashing, they are not worth the effort to make sense of them philosophically. But that too is a mistake. The actions that go into acquiring knowledge-how are not mere thrashing, they are actions that call for a different model to explain their place in a philosophical view of intelligent action and rational agency. I am not going to argue for that view here, but I do want to gesture at a framework for thinking about it.

The right way to approach modeling the acquisition of new intelligent capacities is to think about the actions involved using an analogue of induction. In the theoretical realm, induction involves transitioning from beliefs that are about particulars to beliefs that subsume those particulars under a more general belief. For example, the transition from a belief that ‘this crow is black’, ‘that crow is black’, and so on, to the general belief that ‘all crows are black’ is a transition typical of induction. There is, I think, a similar transition in the case of learning how to do things. From particular instances of doing things, agents transition to more general capacities. The idea, then, is to represent a practical transition from lesser capacities to more general capacities using a model of action that relies on the notion of coordination through a type of agential induction.

Learning to do new things requires learning to coordinate thought with action, but such coordination requires agents to undertake particular activities without knowing what they are doing. For example, to learn how to ski, individuals must perform the particular actions involved in skiing without grasping their relation to the activity as a

whole and without understanding the content of the thoughts that might characterize such activities. The process of coordinating the particular movements involved with skiing with the thoughts that lead to the performance of the action can, of course, eventually lead to the general capacity to ski, but that process, it seems to me, involves a certain type of transition. When an agent has acquired the knowledge-how pertaining to skiing, she has transitioned from particular instances of exercising her agency to a general capacity to do so. If that is right, then a model of action designed to explain the acquisition of knowledge-how should be able to make sense of the idea that learning how to do things is a matter of increasing what one is able to do by coordinating one's thoughts with one's actions.

The idea seems to be anticipated by Aristotle when he writes, “the things we have to learn before we can do, we learn by doing, e.g., men become builders by building and lyre players by playing the lyre” (1984, 1103a30-1103b1). It is by performing particular acts on the lyre—experimenting with different finger positions, plucking strengths, and so on—that agents discover how those acts fit together in a manner consistent with being a lyre player. But of course, prior to the acquisition of that general capacity, individuals do not know how particular actions fit together to effect the skill of playing the lyre. Acquiring the intelligent capacity, then, is characterized by learning, realizing, or discovering how particular actions fit with what one knows by doing things without understanding what one is doing. Only in this way does an individual discover how particular actions fit with thoughts about lyre-playing.

The idea also seems to be anticipated by Ryle when he writes: “It is the essence of intelligent practices that one performance is modified by its predecessors” (1949, 42). Or, again a bit further on, an agent “learns how to do things [by] thinking what he is doing, so that every operation performed is itself a new lesson to him how to perform better” (1949, 43). To see what I have in mind, consider, again, Hannah's attempt to acquire an understanding of how to ski. As she begins, her balance is hit or miss. She catches her edges easily and falls over at the slightest provocation. Sometimes, she can get through three or four turns without losing her balance, but even then, her actions betray an apprehension characteristic of the novice. Since she has not yet coordinated the movements needed to ski with the thoughts required to execute the

skill, she does not really know how to do what she is attempting to do. Nevertheless, with each fall, each turn, and every trip up and down the ski-hill, she slowly gets better, more capable, and more coordinated in her activity. Each performance, as Ryle notes, is a new lesson to her how to perform better.

But what explains the steady improvement of her capacity? Part of the story might be the propositional information she garners from her experience; part of the story might be her inborn capacity to think of that information according to the progressive form of thought. But these pieces of the explanatory story cannot make sense of the things she *does*, the *actions* she undertakes, without knowing what she is doing. To explain an individual's improvement as she learns a new skill, we must account for the things she does without knowing what she is doing. And models of action built with intellectualist assumptions are not designed to do that. Such models cannot account for an obvious component of an agent's acquisition of knowledge-how—namely, the actions performed without the coordinated unity of thought and action. To explain that, we need a different model.

Of course, I believe the right way to think about the design of such models is with an analogue of induction. If agential capacities develop to more general capacities through a process of discovering, learning, or realizing how some particular action fits into some bigger, whole action, we should use the parallel insights of induction to explain an agent's acquisition of intelligent capacities or skills. From less sophisticated actions, an agent is, through the coordination of thought and action, able to extend his or her capacities to more general ones. And it seems to be something analogous to induction that is governing the process. Consequently, it should frame any model designed to explain such actions.

4.7 Conclusion

As embodied creatures, we have to learn how to do nearly everything we know how to do. If practical intellectualism of one variety or another was the right way to think about acquiring knowledge-how, agents would be able to learn how to *F* by merely thinking about what they already know how to do. But such a view is implausible. And it is implausible precisely because we are embodied creatures. We have to learn

to coordinate our thoughts with the actions required for doing the things we do. This coordination of thought and action cannot be learned by reflection merely, nor is it, in a very large range of cases, an inborn capacity. Instead, the acquisition of knowledge-how depends on agents doing things without knowing what they are doing. Yet we lack an explanation of our capacity to act in this manner. To give such an explanation, an account of action needs to be able to explain the things we do without appealing to what an agent already knows. But since nearly all accounts of action explanation assume intellectualism of one variety or another, this means we have to rethink the design parameters for models of action in order to explain this class of action. I have offered what I think might be the right way to do this, but certainly there are other options. Nevertheless, only by developing such alternative models will we be able to get a better grasp of how we learn to be the rational agents that we are.

CHAPTER 5

CONCLUSION

5.1 Introduction

Over the past few chapters, I have been arguing both directly and indirectly for a model-based approach to problems of rational agency. In this concluding chapter, I want to tie the ideas motivating the earlier ones together in an effort to frame some general ideas about the type of agency we possess as human beings. Accomplishing that task will require sketching a bit of background concerning the unity of agency and the widespread assumption that a necessary condition for the possibility of being an agent is being, in some sense, unified. After I have provided that background, I will draw out the implications of the previous chapters for the idea that unified agency is required for intelligent action, and, in the final section, conclude with an argument. To anticipate: to be the type of agents we seem to be requires being disunified in a distinctive sort of way.

5.2 The Role of Unity

That we must act as unified agents is frequently accepted as a starting place for theorizing about human beings as rational agents. The idea seems to be that since individuals have only one body with which to act, doing anything at all requires sufficiently coordinating the various components of one's agency—for example, the desires, plans, preferences, beliefs, and intentions that might cause one to act—so as to avoid, in one sense or another, self-frustration.¹ Since we have only one body with which to act, we have to resolve conflicts between the thoughts and attitudes that might move us to act in order to prevent being pulled in more than one direction. To

¹The idea that we must be unified in one sense or another shows up across the spectrum of philosophical views on agency. See for example Korsgaard (1996; 2008; 2009), Bratman (1987; 2007b), Millgram (1997), Velleman (1989; 2000; 2009), Frankfurt (1998; 1999), and Rödl (2007).

act as a unified agent, then, means coordinating potentially conflicting attitudes and motivations in a way that will produce effective action. There are, however, several different proposals for how this coordination might be effected. Indeed, the disparity between the different proposals for achieving unified agency serve to distinguish a variety of views of rational agency. Nevertheless, the background idea is roughly the same: since acting as an agent requires acting as a unit, acting as a rational agent requires some basis of rational unity.

Given that shared idea, it is worth pausing to reflect on the different ways in which one's agency might be unified. Before I do that, however, we first need to ask, aside from avoiding self-frustration, why is it worth worrying about the unity of agency? Although it is likely that different theorists will emphasize its importance for different reasons, at its heart, the issue seems to me to be one of being in control of the things we do. If an agent fails to act on motives that he endorses, and, instead, acts from motives that are in opposition to what he wants to do, or that are alien to him, the drives moving him will, in one sense or another, be foreign to him as a rational agent. Of course, actions performed from such "alien" drives cannot actually be the agent's actions, because what he is being moved by is divorced from the attitudes he endorses as his own. That is to say that such actions are not actions controlled by the person performing them. One way of overcoming that difficulty is to ensure that what one does is endorsed by thoughts and attitudes essential to an individual's own sense of agency. In other words, one way to overcome that problem is to ensure that one's agency is unified. If that is right, we have some grounds for thinking that unified agency matters.

But we can go further. As human agents, we face a variety of immediate, near-term, and long-term challenges that, unless we are sufficiently in control of the things that move us to act, will be difficult to manage. If our drives get away from us and act through us as alien causes, managing the lives that we lead will be difficult. For example, if I intend to stay in Friday night to get some work done, but fail to coordinate that intention with my immediate desire for some friendly company, I may act in frustration of myself. If I go out, I frustrate the intention to get work done, but if I stay in, I frustrate my desire for company. The problem can be managed by

coordinating my motivational attitudes: I'll stay in and work for a while, and catch up with my friends later. To live with an eye toward the future while living in the here and now, we need to coordinate the components of agency that move us to act. If that is correct, the unity of agency also matters because, in one way or another, it seems required to manage practical challenges that we all face as human agents.

Given the relevance of the idea that we must, to one degree or another, be unified in order to be in control of our lives, we can now consider what we mean when we say that agency must be rationally unified. There are two central ideas to consider: the first is that we need a principle of unity to coordinate actions over time as the actions of a single individual; the second is that we require a principle of unity in order for the things we do on particular occasions to be actions under control. Both notions have shown up during the course of this dissertation. On Bratman's conception of agency, for example, the self-governing policies that hold particular agents together over time, and provide authority for what agents decide to do, must meet certain standards of rationality in order to serve as temporally persisting standpoints of agency. In particular, one's self-governing policies must meet standards of coherence, stability, and consistency. When an agent is satisfied with the self-governing policies that are the backbone of his agential identity, and when such policies meet the relevant standards of rationality, he has a principle of control that can serve to coordinate his actions over time as actions belonging to him. Due to the coordinated unity of his temporally persisting psychological states, a Bratmanian agent can control the things *he* does as actions that reflect his temporally ordered agency. To the extent that such actions are motivated by desires that are in harmony with the agent's higher-order policies, the individual will act as an agent in control of whatever it is that he is doing.

A similar principle of unity, which serves to provide the control necessary for acting as an agent, operates in the background of Velleman's view. For Velleman, an agent takes control of himself when he acts to satisfy the higher-order desire to be in control. The idea, of course, is that by reflecting on what one prefers and desires, and by using beliefs about one's preferences and desires to make inferences about what one actually desires and prefers, agents act in a way that is self-intelligible. Of

course, being intelligible to oneself, and acting on the basis of what one understands about oneself, seems to be required if one is to avoid self-alienation. If an individual did not know why he was doing what he was doing, or if his actions were motivated by attitudes that he failed to recognize as his own, it would be difficult to think of that agent as one in control of himself. Positioning the higher-order desire for control as the point of unity, then, places Velleman in a position to explain both the coordination of action over time and the control necessary for intelligently acting in the here and now.

On views like those endorsed by Bratman and Velleman, the unity of agency depends on bringing into harmony potentially conflicting psychological attitudes. To effect that unity, there must be one or another higher-order conative state that serves as the focal point. That point, of course, is the grounding consideration for rational agency: if an agent's other psychological attitudes fail to operate in the service of that grounding consideration, the agent's rational nature is, to one degree or another, undermined. To be an agent in rational control of one's life, then, individuals must align competing psychological attitudes with higher-order attitudes that serve as the point of unity. When we desire to bring something about, reflectively engage that desire, and, in one way or another, endorse it as a desire we want to be motivationally efficacious, we act as unified agents in control of our lives.

But of course, the type of hierarchical framework endorsed by Velleman and Bratman is only one way to think of the unity of agency. The picture is different when we turn our attention to Anscombian views. For these views, the notion of unity is designed to explain the things we do on particular occasions. How is it that the things we think and the things we do result in actions over which we have control? For Anscombian views, the requisite unity is had through the practical knowledge required for intelligent action. Of course, for theorists that endorse such views, knowing one's way around practical issues—that is, being able to act in a way that displays one's practical knowledge—depends on the formal unity of thought and action. In contrast to instrumentalist views that rely on the psychological unity of individuals to ground rational agency, Anscombian views rely on the idea that control as a single, unified creature depends on the knowledge necessary for accomplishing the

things one has set out to do. Put differently, individuals take control of their actions when they act with an awareness of the relation between what they are doing and what they think they will achieve by doing what they are doing. For theorists of this stripe, the unity of agency depends on the formal link between what an individual knows when acting and the structure of the actions themselves. When what we think we are doing reflects the structure of the actions being performed, we act as individuals in control of our lives. It is the unity between thought and action, and the practical awareness presupposed by that unity, that makes us the agents we are.

Of course, there are other proposed principles of unity in the literature, which I have not dealt with in the course of this dissertation. Christine Korsgaard, for example, argues that acting as an agent depends on resolving tension between disparate psychological proclivities by subjecting them to universalizable principles of reason. Only when the motivational attitudes of an individual are aligned according to such principles (rather than, for example, by the ends an individual wants to achieve or by the shared formal structure of thought and action) are the actions expressed by that unity constitutive of his or her nature as a rational creature. Only in this way, she thinks, can we act as free and rational agents.

In all of these views, certain ideas about the unity of agency, and the structures needed to effect that unity, are central to providing an account of rational agency. Given the presupposition that we must be unified in order to be agents, the problem becomes one of figuring out what the criteria of unification must be. And the thought is that by sorting out what those criteria are, we come to understand what it means to be a rational agent.

5.3 A Sketch of Disunity

I have been arguing that for one reason or another we cannot do without seemingly disparate portraits of rational agency. If I am right, my arguments show that we need different notions of agency to make sense of the diverse aspects of our nature as rational agents. But since these different notions rely on incompatible conceptions of agency, my arguments suggest that we cannot be unified. Indeed, in *Instrumental Facades*, I argued against the idea that a single theory of instrumental agency is

sufficient to address the range of problems human agents have a stake in addressing. Since we all have an interest in managing problems that rely on features of agency that cannot be integrated, but which are sufficiently deep to make abandoning them out of the question, we have to settle for a model-based approach. Due to this fact, there will always be, in some sense, deep fissures between different points of agency that may serve as the unifying basis of our actions. The result is that whenever we do anything, we will, in one way or another, be acting against other aspects of our own agency. In fact, I have argued that it is very unlikely that the deep motivational attitudes at the heart of our agency can be made fully harmonious. Consequently, in acting as instrumentally rational agents, it is unlikely that we will ever fully integrate our conflicting psychological attitudes around a single, unifying point of agency.

Of course, I have also argued that the type of unity required by the Anscombian conception of rational agency is insufficient to make sense of the things we do as intelligent agents. In “Modeling Expressive Actions,” I argued that acts of emotional expression—especially those characteristic of artistic expression—cannot be represented with a view of agency that requires practical knowledge. This, of course, is in contrast to the Anscombian view, which requires thought and action to be unified through practical knowledge in order to act as intelligent creatures. As I demonstrated, painters, poets, dancers and other artists frequently act without knowing what they are doing. But artists are not unique in this respect—that is, there is no reason to think that artists possess some form of agency peculiar only to them. Indeed, common forms of emotional expression seem to be done without practical knowledge. Since how we see the world shifts under the influence of emotion, and since those perceptual shifts influence action in a way that we only understand at best metaphorically, acting from emotion seems to produce actions that are done without the type of practical knowledge required by the Anscombian model. Instead, the things we do to give expression to emotion seem to be tied to thoughts that depend on the structure of emotions themselves. Nevertheless, such actions seem to be intelligently performed. If that is correct, it looks like we can do things as agents without the type of unity required by Anscombian views of action explanation. Consequently, we should think anew about the notion of unified agency and what it

means to express emotion through action.

A similar conclusion fell out of Chapter 4. There I argued that learning to coordinate thought and action in a manner that results in knowledge-how requires acting without knowing what one is doing. Such actions, of course, violate the principle of unity behind the Anscombian view of intelligent action. As with expressive acts, when we are learning to perform actions that we have never before performed, we lack the practical knowledge that, on the Anscombian view, is required to explain intelligent action. Indeed, if it is true that we must perform actions without knowing what we are doing to acquire knowledge-how, the Anscombian idea that all intelligent actions require acting with practical knowledge cannot be right. Instead, there must be actions that are performed on purpose and for reasons without the tight connection characteristic of the model endorsed by Rödl. In fact, such actions must be governed by some other way of thinking, which I suggested should be modeled with an analogue of induction. The idea depends on the fact that there are things we do, and do intelligently, that cannot be modeled with the type of unity that is thought to explain the actions of rational agents. As with expressive actions, then, actions performed in the course of acquiring new capacities seem to undermine the idea that a single type of formal unity is necessary for acting as intelligent agents.

These later chapters, which show that the Anscombian view needs to be reconsidered, should not be confused as arguments to the effect that some of what we do is arational, unintelligent, or unintentional. In fact, the alternative modes of action I have tried to characterize in these chapters are intended to show that some of the things we do as intelligent creatures—that is, some of our thoughtful actions—must be represented with different models of action explanation. When an artist gives expression to his emotions through painting, he is acting as an individual in control of himself, but the thought governing the actions being performed is not best represented with traditional philosophical views. Similarly, when an individual acquires new intelligent capacities by acting without practical knowledge, what he does seems to be done with control and as an intelligent action. Yet current models of action explanation fail to capture that fact. These points suggest that the types of action for which I have argued in these later chapters show different ways in which

our agency can be tied to the things we do. The principle of control according to which we express our emotions, and the principle of control behind the acquisition of new intelligent capacities, need to be accounted for with different models of action explanation. But again, that does not suggest that such actions are, in some sense, unintelligent. To the contrary, it suggests that there is more than one way for us to act as intelligent creatures.

Of course, from all of this, we cannot conclude that there is *no* overarching structure that can serve to unify these seemingly disparate ways of acting as agents, but I do think it gives us reason to be skeptical. In particular, given the fact that some forms of action seem to be governed by thoughts that do not require practical knowledge, it is unlikely that our actions as intelligent agents will require doing things according to universalizable, self-determined laws of reason. After all, to prescribe a law of action to oneself, which applies to all rational agents as such, one must be aware of it as a law. But if some of what we do as intelligent creatures is done without the type of practical self-knowledge required by the views we have been surveying, it seems implausible to think that those actions could be governed by such self-determined laws. We seem, then, to have some basis for being skeptical of accounts that say otherwise.

5.4 Models of Agency

Before wrapping things up, let me pause for a moment. What I have been suggesting over the course of this final chapter is that there are different ways to understand intelligent action. That fact suggests that to the extent that we act intelligently in these different ways, there are different senses to be given to the idea of unity of agency. Indeed, I believe that, for one reason or another, all of these different senses of being unified are needed to make sense of ourselves as the complex creatures we are. The idea depends on understanding the different notions of agency as patches in a complex array of disjoint, but still connected, ideas about what it means to be a rational agent. For instrumentalist views like those of Bratman and Velleman, understanding agency depends on locating motivational features at the heart of our psychology and using them as focal points of integration. To be a

rational agent is for one's lower-order attitudes to be integrated to the attitudes that are, in some sense, necessary for agency. For Anscombian views, unity depends on bringing thought and action together according to a specific formal structure. To be an agent in this sense is for one's actions to unfold in the manner typified by one's practical thought. Of course, for the alternative type of actions for which I have argued, being an agent turns out to be a matter of either expressing oneself according to the structure of emotions or on the basis of practical thoughts that should be modeled with an analogue of induction. Actions performed on the basis of either of these ways of thinking depend on a type of unity with which philosophers are unfamiliar.

What we have before us, then, is a collection of ways of being an agent that, when combined, suggests a picture of ourselves as agents that is incongruous. Instead, we are left with a patchwork of representations that are only loosely connected in the sense that they are all part of an array of concepts necessary for understanding ourselves as rational human agents. To the extent that we express our emotions in an intelligent manner without practical knowledge, that feature of our agency cannot be integrated with the type of intentional actions characteristic of Anscombian views. So too, because acquiring knowledge-how depends on acting without knowing what we are doing, and because such actions are more than mere behavioral thrashing, they too cannot be integrated with the Anscombian view. Nevertheless, some of what we do should be modeled using the type of structure found in Anscombian models of action explanation as well. There is a similar point to be made about the models surveyed in Chapter 2. To make sense of aspects of our agency that seem important for being the type of agents we are, we need views of instrumental agency that are at odds with each other. The result, however, is that there is a similar sort of hinderance to the integration of deep features of our motivational psychology. All of these points together suggest that, even if we can do different things by being sufficiently unified in different ways, we cannot be fully integrated as human agents. To be disunified in this distinctive sort of way appears to be a consequence of the type of agents we are.

Of course, if we cannot in fact be fully unified, then to make sense of ourselves as creatures capable of intelligent action, we are going to need a variety of representations

to capture the different ways in which we act intelligently. Models seem like a natural fit: since no single representation will do, a patchwork of models that allow for a variety of representations seems like the right way to figure out what it means to be a rational human agent. I have tried to show over the course of this dissertation how such a patchwork of models might begin to look, and to provide a clearer picture of the ways in which we act as creatures capable of moving ourselves for reasons.

REFERENCES

- Anscombe, Elizabeth. 1995. "Practical Inference." In *Virtues and Reasons*, eds. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, 1–34. Oxford: Clarendon Press.
- Anscombe, G.E.M. 2000. *Intention*. Cambridge: Harvard University Press.
- Aristotle. 1984. "Nicomachean Ethics." In *The Complete Works of Aristotle*, ed. Jonathan Barnes, volume 2, 1729–1867. Princeton: Princeton University Press.
- Austin, J.L. 1961. "A Plea for Excuses." In *Philosophical Papers*, eds. J.O. Urmson and G.J. Warnock, 123–152. Oxford: Oxford University Press.
- Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge: Harvard University Press.
- . 2007a. "Reflection, Planning, and Temporally Extended Agency." In *Structures of Agency*, 21–46. Oxford: Oxford University Press.
- . 2007b. *Structures of Agency*. Oxford: Oxford University Press.
- . 2007c. "Three Theories of Self-Governance." In *Structures of Agency*, 222–253. Oxford: Oxford University Press.
- Brooks, Cleanth and Warren, Robert. 1976. *Understanding Poetry*. New York: Holt, Rinehart and Winston, 4th edition.
- Carnap, Rudolf. 1939. *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press.
- Carroll, Lewis. 1895. "What the Tortoise Said to Achilles." *Mind* 4:278–280.
- Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.
- . 1999. "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Model of Superconductivity." In *Models as Mediators*, eds. Mary Morgan and Margaret Morrison, 241–281. Cambridge: Cambridge University Press.
- Collingwood, Robin George. 1958. *The Principles of Art*. Oxford: Oxford University Press.
- Cooper, John M. (ed.). 1997. *Plato: Complete Works*. Indianapolis: Hackett Publishing Company.

- Damasio, Antonio. 1994. *Descartes' Error*. New York: Penguin Books.
- Davidson, Donald. 1980. "Actions, Reasons, and Causes." In *Essays on Action and Events*, 3–19. Oxford: Clarendon Press.
- de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge: The MIT Press.
- Dewey, John. 1934. *Art as Experience*. New York: Capricorn Books.
- Downes, Steve. 1992. "The Importance of Models in Theorizing: A Deflationary Semantic View." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, eds. D. Hull, M. Forbes, and K Okruhlik, volume 1, 142–153. East Lansing, MI: Philosophy of Science Association.
- Ford, Anton, Hornsby, Jennifer, and Stoutland, Frederick (eds.). 2011. *Essays on Anscombe's Intention*. Cambridge: Harvard University Press.
- Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68:5–20.
- . 1988. "The Problem of Action." In *The Importance of What We Care About*, 69–79. Cambridge: Cambridge University Press.
- . 1992. "The Faintest Passion." *Proceedings and Addresses of the American Philosophical Association* 66:5–16.
- . 1998. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- . 1999. *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- Giere, Ronald. 1988. *Explaining Science*. Chicago: The University of Chicago Press.
- Goldie, Peter. 2010. *The Oxford Handbook of Philosophy of Emotion*. Oxford: Oxford University Press.
- Hempel, Carl G. 1965. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press.
- Hospers, John. 1954–1955. "The Concept of Artistic Expression." *Proceedings of the Aristotelian Society* 55:313–344.
- Hursthouse, Rosalind. 1991. "Arational Actions." *The Journal of Philosophy* 88:57–68.

- Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32:127–136.
- Kant, Immanuel. 1997. *Groundwork of the Metaphysics of Morals*. Cambridge: Translated by Mary Gregor. Cambridge University Press.
- Korsgaard, Christine. 1996. "Personal Identity and the Unity of Agency." In *Creating the Kingdom of Ends*, 363–397. Cambridge: Cambridge University Press.
- . 2001. "Skepticism About Practical Reason." In *Varieties of Practical Reason*, ed. Elijah Millgram, 104–125. Cambridge: MIT Press.
- . 2008. *The Constitution of Agency*. Oxford: Oxford University Press.
- . 2009. *Self-Constitution*. Oxford: Oxford University Press.
- Millgram, Elijah. 1996. "Williams' Argument Against External Reasons." *Nous* 30:197–220.
- . 1997. *Practical Induction*. Cambridge: Harvard University Press.
- Morrison, Margaret. 1999. "Models as Autonomous Agents." In *Models as Mediators*, eds. Mary Morgan and Margaret Morrison, 38–65. Cambridge: Cambridge University Press.
- Nussbaum, Martha. 1988. "Narrative Emotions: Beckett's Genealogy of Love." *Ethics* 98:225–254.
- O'Neill, Onora. 2001. "Universality Tests in Autonomous and in Heteronomous Ethics." In *Varieties of Practical Reason*, ed. Elijah Millgram, 301–329. Cambridge: MIT Press.
- Pavese, C. and Stanley, J. 2011. "Speaking of Practical Knowledge." *Proceedings of Semantics and Linguistic Theory* 20.
- Plato. 1997a. "Meno." In *Plato: Complete Works*, ed. John M. Cooper, 870–897. Indianapolis: Hackett Publishing Company.
- . 1997b. "Republic." In *Plato: Complete Works*, ed. John M. Cooper, 971–1223. Indianapolis: Hackett Publishing Company.
- Ravetz, Jerome R. 1971. *Scientific Knowledge and its Social Problems*. Oxford: Oxford University Press.
- Rödl, Sebastian. 2007. *Self-Consciousness*. Cambridge: Harvard University Press.

- . 2011. “Two Forms of Practical Knowledge and Their Unity.” In *Essays on Anscombe’s Intention*, eds. Anton Ford, Jennifer Hornsby, and Frederick Stoutland, 211–241. Cambridge: Harvard University Press.
- Rorty, Amelie (ed.). 1980. *Explaining Emotions*. Berkeley: University of California Press.
- Ryle, Gilbert. 1949. *The Concept of Mind*. New York: Barnes and Noble.
- Setiya, Kieran. 2008. “Practical Knowledge.” *Ethics* 118:388–409.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sparshott, Francis. 1997. “Emotion and Emotions in Theatre Dance.” In *Emotion and the Arts*, eds. Mette Hjort and Sue Laver, 119–136. Oxford: Oxford University Press.
- Spicer, Jack. 1998. *The House That Jack Built: The Collected Lectures of Jack Spicer*. Hanover, CT: Edited by Peter Gizzi. Wesleyan University Press.
- Stanley, Jason. 2011a. *Know How*. Oxford: Oxford University Press.
- . 2011b. “Knowing (How).” *Nous* 45:207–238.
- Stanley, Jason and Williamson, Timothy. 2001. “Knowing How.” *The Journal of Philosophy* 98:411–444.
- Suppes, Frederick. 1977. *The Structure of Scientific Theories*. Champaign: University of Illinois Press, 2nd edition.
- Suppes, Patrick. 1960. “A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences.” *Synthese* 12:287–301.
- Sylvester, David. 1980. *Interviews With Francis Bacon*. London: Thames & Hudson.
- . 2001. *Interviews with American Artists*. New Haven: Yale University Press.
- Thompson, Michael. 2008. *Life and Action*. Cambridge: Harvard University Press.
- van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Velleman, David. 1989. *Practical Reflection*. Princeton: Princeton University Press.
- . 2002. “Identification and Identity.” In *Contours of Agency*, eds. Sarah Buss and Lee Overton, 91–123. Cambridge: MIT Press.

- . 2009. *How We Get Along*. Cambridge: Cambridge University Press.
- . 2000. *The Possibility of Practical Reason*. Oxford: Clarendon.
- Vogler, Candace. 2002. *Reasonably Vicious*. Cambridge: Harvard University Press.
- Watson, Gary. 2004. "Free Agency." In *Agency and Answerability*, 13–32. Oxford: Oxford University Press.
- Williams, Bernard. 1981a. "Internal and External Reasons." In *Moral Luck*, 101–113. Cambridge: Cambridge University Press.
- Williams, Bernard. 1981b. *Moral Luck*. Cambridge: Cambridge University Press.
- Wilson, Mark. 2006. *Wandering Significance*. New York: Oxford University Press.